languages

Recent Advances in Research on Island Phenomena

Edited by

Anne Mette Nyvad and Ken Ramshøj Christensen

Printed Edition of the Special Issue Published in *Languages*

www.mdpi.com/journal/languages

MDPI

Recent Advances in Research on Island Phenomena

Recent Advances in Research on Island Phenomena

Editors

Anne Mette Nyvad
Ken Ramshøj Christensen

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editors

Anne Mette Nyvad
Aarhus University
Denmark

Ken Ramshøj Christensen
Aarhus University
Denmark

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Languages* (ISSN 2226-471X) (available at: https://www.mdpi.com/journal/languages/special_issues/Recent_Advances_in_Research_on_Island_Phenomena).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-6316-9 (Hbk)
ISBN 978-3-0365-6317-6 (PDF)

Cover image courtesy of Ken Ramshøj Christensen.

Contents

Preface to "Recent Advances in Research on Island Phenomena"

This book consists of a collection of papers presenting state-of-the-art research on the phenomenon known as 'syntactic islands'—structural configurations that to varying degrees resist word order variation involving extraction. The focus is on data from English and the Scandinavian languages, and the volume presents accounts from different theoretical perspectives (including minimalist syntax and construction grammar) taking different methodological approaches (e.g., corpus linguistics, experimental linguistics, and theoretical syntax).

Anne Mette Nyvad and Ken Ramshøj Christensen
Editors

languages

MDPI

Editorial

Recent Advances in Research on Island Phenomena

Anne Mette Nyvad * and Ken Ramshøj Christensen *

Department of English, School of Communication and Culture, Aarhus University, 8000 Aarhus C, Denmark
* Correspondence: amn@cc.au.dk (A.M.N.); krc@cc.au.dk (K.R.C.)

In natural languages, syntactic elements can, in principle, be linked across an unbounded distance, as exemplified by filler-gap dependencies (also known as extractions or movement operations). However, while distance is in itself not a constraint, there are a number of structures or constructions that to varying degrees hinder such long-distance dependencies. Since Ross (1967), the term "island" has been used to describe syntactic structures from which extraction is impossible or impeded, and the constraints on such dependencies have typically been assumed to be universal properties of language and to be innate, given the lack of negative evidence during language acquisition (see, e.g., Newmeyer 1991).

English has been the prototypical object of study in accounts trying to establish what is possible and impossible with respect to long-distance filler-gap dependencies, e.g., across clausal boundaries, and it is still ubiquitous in the literature on island structures. However, counterexamples in the Mainland Scandinavian languages were first brought to the attention of the linguistic community as early as the 1970s (Erteschik-Shir 1973; Engdahl and Ejerhed 1982), and had in fact been recorded by the Danish grammarian Kristian Mikkelsen almost a century before (Mikkelsen 1894, e.g., pp. 322, 442). Nonetheless, such counterexamples have repeatedly been dismissed as illusory and, instead, alternative accounts of the underlying structure of such cases have been proposed (e.g., Chomsky 1982; Kush et al. 2013).

In fact, extractions from island structures appear to be pervasive in spoken Mainland Scandinavian (Lindahl 2017; Nyvad et al. 2017), and recent experimental island research on Swedish (Müller 2019), Norwegian (Kush et al. 2018, 2019; Bondevik et al. 2020) and Danish (Christensen and Nyvad 2014) have found empirical evidence to suggest that adjunct clauses and relative clauses may not actually be strong (absolute) islands in these languages. This recent island research has highlighted the roles played by the matrix verb, information structure, and the faciliatory effect of a supporting context in the acceptability of extraction from island environments. In addition, recent research suggests that extraction from certain types of island structures in English might not be as unacceptable as previously assumed either (e.g., Müller 2019; Chaves and Putnam 2020; Sprouse et al. 2016). These findings break new empirical ground, question perceived knowledge, and may indeed have substantial ramifications for syntactic theory.

The purpose of this Special Issue is to provide an overview of the state-of-the-art research on island phenomena in English and the Mainland Scandinavian languages, as well any other languages where such island structures can be found. An explicit objective is to investigate how other languages compare to English with respect to the acceptability of extractions from island structures in order to shed light on the nature of the constraints on filler-gap dependencies and the syntactic primitives that form the basis of such structures.

Each of the ten contributions of this Special Issue play a part in highlighting the need to operationalize our tools for the investigation into the extra-syntactic factors that affect the acceptability of traditional island structures. However, each of the ten contributions approach island phenomena from a distinctive angle. While some take a more theoretical approach (Culicover et al. 2022; Kehl 2022), others provide new corpus data (Müller and

Citation: Nyvad, Anne Mette, and Ken Ramshøj Christensen. 2023. Recent Advances in Research on Island Phenomena. *Languages* 8: 16. https://doi.org/10.3390/languages8010016

Received: 21 December 2022
Accepted: 22 December 2022
Published: 2 January 2023

Eggers 2022; Engdahl and Lindahl 2022) or new experimental data (Chaves 2022; Engdahl and Lindahl 2022; Kobzeva et al. 2022; Nyvad et al. 2022; Snyder 2022; Vincent et al. 2022).

In order to investigate the nature of island constraints in the Mainland Scandinavian languages, the Insular Scandinavian languages could be of special interest, since these languages share a number of typological and syntactic features, but appear to be different when it comes to the escapability of island environments. Along this line of reasoning, Engdahl and Lindahl (2022) argue that the extractions from relative clauses attested in the Mainland Scandinavian languages could conceivably be connected to the tendency in these languages to prepose unstressed object pronouns as a way of establishing coherent discourse. In order to test this hypothesis, they examine the extent to which pronoun preposing is employed in the Mainland Scandinavian language Swedish on the one hand, and the Insular Scandinavian languages Faroese and Icelandic on the other. Their findings suggest a connection between the frequency with which a language employs pronoun preposing in order to establish coherent discourse and the willingness of its speakers to accept long extractions. In addition, some Icelandic speakers appear to find extraction from complement clauses natural, but extraction from relative clauses (RC) unnatural, which suggests that there is a syntactic constraint blocking movement from RCs in Icelandic. While the Mainland Scandinavian languages seem to share a property facilitating long extraction, corpus data suggest that such extraction is not used productively in the Insular Scandinavian languages.

Building on these new insights about fronting conditions in Mainland and Insular Scandinavian languages, Lindahl (2022) examines extraction from relative clauses in Icelandic and Swedish in parallel with an acceptability judgment study in order to compare island structures when information structure and context are controlled for. She found substantial differences between the two languages: Participants judged extraction from relative clauses and embedded questions unacceptable in Icelandic, while these island structures were largely considered natural by Swedish speakers. In her data, there appears to be no cost of extraction in local fronting, extraction from complement clauses, or from 'existential/presentational' RCs, i.e., RCs embedded under existential or presentational main verbs, corresponding to, e.g., *be* or *know*. While extractions from RCs under semantically richer matrix predicates were generally considered less natural in Swedish, extraction from RCs was invariably unacceptable in Icelandic.

However, there is often a mismatch between informal and formal judgments. Island structures tend to receive relatively low scores in acceptability judgment studies, while highly acceptable examples are frequently attested in naturally occurring speech. As already mentioned, speakers of Mainland Scandinavian languages find at least some extractions from relative clauses and adjuncts acceptable, and the finding in Müller and Eggers (2022) is that this is matched in naturally occurring speech: island violations can be attested in corpora "at a non-trivial rate". The combined data from corpus and acceptability studies strongly suggest that extraction from relative clauses and adjunct clauses is possible in Danish, or at least that they are not the strong islands that they have traditionally been assumed to be. However, Müller and Eggers also found examples of extraction from RCs and adjunct clauses in English, albeit on a substantially smaller scale. In other words, the data suggest that the Mainland Scandinavian languages are more liberal than, e.g., English when it comes to island extraction, and that RCs and adjunct clauses may not be strong islands in English either. The data also show that extraction from RCs is not restricted to constructions involving existential/presentational matrix verbs, although there seems to be a strong preference (echoing Lindahl 2022). In addition, while relativization from RCs and adjunct clauses was attested in both English and Danish, none of the examples involve *wh*-extraction.

In Müller and Eggers (2022), extraction from adjunct clauses was most frequent from *if*-clauses and unattested from *because*-clauses in both Danish and English. These results echo the experimental results by Nyvad et al. (2022), who investigated the acceptability of relativization from three types of English finite adjunct clauses (headed by *if*, *when* and

because) with a facilitating context and found a quite non-uniform pattern: extraction from *if*-clauses received unexpectedly high ratings, on a par with extraction from complement *that*-clauses, and significantly higher than extraction from *when*- and *because*-clauses. The authors interpret these results as suggesting that adjunct clauses are not invariably strong islands in English and that extra-syntactic factors must be key in understanding the variation in these island phenomena.

Similarly, Vincent et al. (2022) argue that English is more like the Mainland Scandinavian languages than previously thought. It has been suggested that island constraints need to be parameterized in order to account for cross-linguistic differences, but a substantial part of the research presented in this Special Issue suggests that island extraction appears to be facilitated in similar environments across the languages under investigation. This conclusion is also reached in Vincent et al. (2022), based on experimental data indicating that extraction from a relative clause is possible or at least more acceptable in English in environments and contexts where an existential, non-presuppositional interpretation of the DP containing the relative clause is supported.

Kobzeva et al. (2022) examine extractability in Norwegian. They used an acceptability judgment task to test relativization and *wh*-extraction from nominal subjects, embedded questions, conditional adjunct clauses, and existential relative clauses in Norwegian. The results reveal different effects for different dependency types across the island environments, as *wh*-extraction from embedded questions and conditional adjunct clauses induce small but significant island effects, while the corresponding relativizations do not. They interpret the results as calling for a fine-grained account of factors influencing the acceptability of island extractions that goes beyond the basic division between focus and background, as suggested by Abeillé et al. (2020).

Culicover et al. (2022) propose the Extended Radical Unacceptability Hypothesis (ERUH) for extraction phenomena and argue that the reduced acceptability found in locally well-formed structures are not created by violations of syntactic constraints, but rather by non-syntactic factors. They assume that prior exposure leads to the emergence of probabilistic expectations that can be described as symbolic local well-formedness conditions (LWFCs), and that the level of acceptability is related to the degree of surprisal triggered by a linguistic form, such that low surprisal corresponds to high acceptability and vice versa. Hence, they argue, the unacceptability of classical island structures may just be a reflection of the level of surprisal that they yield, and what has been described as syntactic island constraints are in fact simply generalizations concerning the structures that produce a high level of surprisal related to their frequency. The frequency with which a given construction occurs is affected by both processing factors and information structure mismatches. A problem for non-syntactic accounts is the cross-linguistic variation that we see in terms of island strength. Culicover et al. suggest that the variation is the result of differences in construction frequencies.

In a similar vein, Kehl (2022) argues that it may not be necessary to make syntactic operations sensitive to semantic factors in accounts of participial adjunct islands. According to him, theory development in the realm of syntax should take into account relative acceptability differences in the underlying declaratives before positing licensing mechanisms for interrogative island structures. The declarative structures have traditionally been viewed as grammatical in a binary, categorical sense, and as such, "[d]ifferences in processing complexity, semantic compatibility and pragmatic characteristics" of the declaratives have not been taken into account in the comparison with the interrogative counterparts when it comes to island structures, but acceptability differences in the declarative constructions might explain at least some of the variation found in the extraction counterparts. Kehl argues that among the factors that may play a role in acceptability variation are transitivity, event structure of the main verb, and the encoding of an incremental measure scale in the matrix predicate.

Snyder (2000) presented experimental evidence that syntactic satiation can be induced for certain island structures. That is, speakers find such sentences more acceptable as a

function of repeated exposure. This is notable, given that satiation effects can potentially help shed light on the fundamental questions relating to the nature of island constraints in particular and the nature of our internal grammars in general. In this Special Issue, Snyder (2022) presents three new studies indicating (i) that only a subset of sentence types satiates, (ii) that the satiation of one sentence type may affect other syntactically related sentence types, and (iii) that satiable sentence types vary in terms of the number of exposures required in order for acceptability to increase. Finally, it is shown that experimentally induced satiation may persist over a certain period of time. Snyder (2022) concludes by suggesting that satiation can be used as a diagnostic test in that the underlying principle leading to initial unacceptability in different sentence types may not be the same if they differ in their behavior in terms of satiability.

Chaves (2022) points out that it is still unclear what the precise nature of syntactic satiation is, i.e., whether it is task adaptation, syntactic adaptation, or both. He stresses that certain island structures may combine categorical (competence) effects and contextual or expectation-based (performance) factors, which may be difficult to separate. However, across studies, it has been found that coordinate structure violations, subject islands, adjunct islands, factive islands, and RCs share a common property in being more acceptable when the construction expresses an assertion, rather than backgrounded or non-at-issue content. In addition to complexity and plausibility as important factors in gradient acceptability, Chaves (2022) presents a new factor that may also play a role in the adaptation to islands and other complex constructions, namely "predicted reward" (i.e., some sort of bonus for accurate performance in an experiment). Based on new experimental data, he suggests that the variation found when it comes to research on satiation effects may be due to differences in motivation and focus among the participants, something which may be manipulated in future research on islands.

In conclusion, based on mounting evidence, and not least the data and analyses presented in this Special Issue, English and the Mainland Scandinavian languages may be more similar when it comes to island phenomena than previously assumed. Across these languages, extraction seems to increase in acceptability (to varying degrees) when the matrix predicate is existential or presentational, when facilitated by context, when the embedded structure is asserted, or when the dependency type is relativization (or topicalization) rather than *wh*-movement. Furthermore, syntactic constraints on extraction, such as the Condition on Extraction Domain (CED, Huang 1982) and the Complex NP Constraint (CNPC, Ross 1967) may need to be re-evaluated, as recent studies have found a wide variation in acceptability that may not prima facie be easily compatible with a binary constraint in core syntax. While little doubt remains that it is more difficult to extract from some domains than others, whether these patterns are the result of competence or performance factors is still an open question.

Author Contributions: Writing—orginal draft, A.M.N. Writing—review and editing, A.M.N. and K.R.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Independent Research Fund Denmark (grant ID: DFF-9062-00047B).

Conflicts of Interest: The authors declare no conflict of interest.

References

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from Subjects: Differences in Acceptability Depend on the Discourse Function of the Construction. *Cognition* 204: 104293. [CrossRef] [PubMed]

Bondevik, Ingrid, Dave Kush, and Terje Lohndal. 2020. Variation in Adjunct Islands: The Case of Norwegian. *Nordic Journal of Linguistics* 44: 223–54. [CrossRef]

Chaves, Rui P. 2022. Sources of Discreteness and Gradience in Island Effects. *Languages* 7: 245. [CrossRef]

Chaves, Rui P., and Michael T. Putnam. 2020. *Unbounded Dependency Constructions: Theoretical and Experimental Perspectives*. Oxford: Oxford University Press.

Chomsky, Noam. 1982. Some concepts and consequences of the theory of government and binding. In *Linguistic Inquiry Monographs*. Cambridge: MIT Press, p. 6.

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2014. On the Nature of Escapable Relative Islands. *Nordic Journal of Linguistics* 37: 29–45. [CrossRef]
Culicover, Peter W., Giuseppe Varaschin, and Susanne Winkler. 2022. The Radical Unacceptability Hypothesis: Accounting for Unacceptability without Universal Constraints. *Languages* 7: 96. [CrossRef]
Engdahl, Elisabet, and Eva Ejerhed, eds. 1982. *Readings on Unbounded Dependencies in Scandinavian Languages*. Acta Universitatis Umensis 43. Umeå and Stockholm: Universitetet i Umeå, Almqvist & Wiksell International [distributor].
Engdahl, Elisabet, and Filippa Lindahl. 2022. Extraction and Pronoun Preposing in Scandinavian. *Languages* 7: 128. [CrossRef]
Erteschik-Shir, Nomi. 1973. On the Nature of Island Constraints. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
Huang, Cheng-Teh James. 1982. Logical Relations in Chinese and the Theory of Grammar. Ph.D. dissertation, MIT, Cambridge, MA, USA.
Kehl, Andreas. 2022. Extraction from Present Participle Adjuncts: The Relevance of the Corresponding Declaratives. *Languages* 7: 177. [CrossRef]
Kobzeva, Anastasia, Charlotte Sant, Parker T. Robbins, Myrte Vos, Terje Lohndal, and Dave Kush. 2022. Comparing Island Effects for Different Dependency Types in Norwegian. *Languages* 7: 197. [CrossRef]
Kush, Dave, Akira Omaki, and Norbert Hornstein. 2013. Microvariation in Islands? In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambruidge: Cambridge University Press, pp. 239–265.
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating Variation in Island Effects: A Case Study of Norwegian Wh-Extraction. *Natural Language & Linguistic Theory* 36: 743–79. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the Island Sensitivity of Topicalization in Norwegian: An Experimental Investigation. *Language* 95: 393–420. [CrossRef]
Lindahl, Filippa. 2017. *Extraction from Relative Clauses in Swedish*. Gothenburg: University of Gothenburg.
Lindahl, Filippa. 2022. Extraction from Relative Clauses in Icelandic and Swedish: A Parallel Investigation. *Languages* 7: 163. [CrossRef]
Mikkelsen, Kristian. 1894. *Dansk Sproglære, Med Sproghistoriske Tillæg: Haandbog for Lærere Og Viderekomne*. Copenhagen: Lehman & Stages Forlag.
Müller, Christiane. 2019. *Permeable Islands. A Contrastive Study of Swedish and English Adjunct Clause Extractions*. Lund: Lund University.
Müller, Christiane, and Clara Ulrich Eggers. 2022. Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English. *Languages* 7: 125. [CrossRef]
Newmeyer, Frederick J. 1991. Functional Explanation in Linguistics and the Origins of Language. *Language & Communication* 11: 3–28. [CrossRef]
Nyvad, Anne Mette, Christiane Müller, and Ken Ramshøj Christensen. 2022. Too True to Be Good? The Non-Uniformity of Extraction from Adjunct Clauses in English. *Languages* 7: 244. [CrossRef]
Nyvad, Anne Mette, Ken Ramshøj Christensen, and Sten Vikner. 2017. CP-Recursion in Danish: A *c*P/CP-Analysis. *The Linguistic Review* 34: 449–77. [CrossRef]
Ross, John Robert. 1967. Constraints on Variables in Syntax. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
Snyder, William. 2000. An Experimental Investigation of Syntactic Satiation Effects. *Linguistic Inquiry* 31: 575–82. [CrossRef]
Snyder, William. 2022. On the Nature of Syntactic Satiation. *Languages* 7: 38. [CrossRef]
Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental Syntax and the Variation of Island Effects in English and Italian. *Natural Language & Linguistic Theory* 34: 307–44. [CrossRef]
Vincent, Jake W., Ivy Sichel, and Matthew W. Wagers. 2022. Extraction from English RCs and Cross-Linguistic Similarities in the Environments That Facilitate Extraction. *Languages* 7: 117. [CrossRef]

languages

MDPI

Article

Extraction and Pronoun Preposing in Scandinavian

Elisabet Engdahl [1,*] and Filippa Lindahl [1,2]

[1] Department of Swedish, University of Gothenburg, Box 200, 405 30 Gothenburg, Sweden
[2] Division of Educational Science and Languages, University West, 461 86 Trollhättan, Sweden; filippa.lindahl@hv.se
* Correspondence: elisabet.engdahl@svenska.gu.se

Abstract: It has been noted that examples with extractions out of relative clauses that have been attested in Danish, Norwegian and Swedish are judged to be unacceptable in Icelandic and Faroese. We hypothesize that this may reflect whether or not speakers tend to prepose unstressed object pronouns as a way of establishing a coherent discourse. In this article we investigate to what extent pronoun preposing is used in Swedish, Icelandic and Faroese and whether there is any correlation with the acceptabilty of extractions from relative clauses. We show that Icelandic speakers use pronoun preposing to a very limited extent whereas Faroese speakers often prepose the VP or sentential anaphor *það*. In both languages extraction from relative clauses is mainly judged to be unacceptable, with Faroese speakers being somewhat more accepting of extraction from presentational relatives. A crucial factor seems to be whether preposing is associated with a marked, contrastive interpretation or not.

Keywords: preposing; topicalization; Faroese; Icelandic; Swedish; contrastive topic; continued topic; VP ellipsis

Citation: Engdahl, Elisabet, and Filippa Lindahl. 2022. Extraction and Pronoun Preposing in Scandinavian. *Languages* 7: 128. https://doi.org/10.3390/languages7020128

Academic Editors: Anne Mette Nyvad and Ken Ramshøj Christensen

Received: 4 March 2022
Accepted: 10 May 2022
Published: 23 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

At least since Ross' dissertation (Ross 1967), linguists have been trying to characterize the restrictions on extractions in natural languages. Ross showed that many types of extractions could be subsumed under a structural constraint, the so called *Complex NP Constraint* CNPC, which prohibited extraction out of a clause dominated by a noun phrase. The CNPC was soon subsumed under *subjacency* (Chomsky 1973) and in the following years, a number of proposals were made for what the relevant bounding nodes were in different languages in order to account for the cross-linguistic variation that was found (see, e.g., Rizzi 1990; Taraldsen 1981). However, soon Scandinavian speaking linguists such as Erteschik-Shir (1973); Andersson (1975) and Allwood (1976) pointed out that certain extractions out of relative clauses were possible in the mainland Scandinavian languages Danish, Norwegian and Swedish. In the following years several linguists approached this topic from different perspectives (see, e.g., the papers in Engdahl and Ejerhed 1982). When written and spoken corpora became available, studies based on spontaneously produced written or spoken extractions started to appear see, e.g., (Engdahl 1997; Jensen 2002; Lindahl 2010, 2011, 2017b).

The documented spontaneous extractions mainly come from mainland Scandinavian whereas the insular Scandinavian languages Faroese and Icelandic appear to behave more similar to English when it comes to extractions from relative clauses (Platzack 2014; Thráinsson 2007; Thráinsson et al. 2004; Zaenen 1985). There have been occasional reports of spontaneous extractions from relative clauses in English (see, e.g., Chung and McCloskey 1983; McCawley 1981)[1] and similar examples are reported by Cinque (2013) to be at least marginally acceptable in Italian, French, and Spanish. However, they do not seem to be used productively the way they are in mainland Scandinavian, see the recent corpus investigations in Kush et al. (2021) and Müller and Eggers (2022). This suggests that the mainland Scandinavian languages have some common property which facilitates long extractions. In this article we explore the hypothesis that the common property that sets

the mainland Scandinavian languages apart is the tendency to use *preposing* (*topicalization*) of unstressed pronouns as a way of connecting utterances in spoken language as well as sentences in texts.[2] We test this hypothesis by comparing how pronoun preposing is used in spoken Swedish, a mainland Scandinavian language, with similar data from the insular languages Icelandic and Faroese. In this article we bring together results from previous studies that we have carried out on individual languages and apply an explicit comparative perspective.

We start in Section 2 by looking at what characterizes spontaneous extractions from relative clauses in Swedish. Since a large proportion of them involve pronouns, we turn, in Section 3, to the question how pronoun preposing is used more generally in Swedish and what similarities there are between local preposing (within a clause), long preposing (from an embedded clause) and what is often called extraction (long preposing from a relative clause). In Section 4, we look at preposing and extraction in Icelandic and in Section 5 we present relevant data from Faroese. In Section 6, we look briefly at the ways preposing is used in English, German and Dutch before evaluating our hypothesis in Section 7.

2. What Do Spontaneous Extractions Look Like?

The literature on extractions from relative clauses is largely based on constructed examples—not surprising given that most of the studies deal with languages where such extractions are not used in ordinary conversations. It is clearly a good idea to start from the types of extractions from relative clauses that are actually used. Engdahl (1997) looked at 30 naturally occurring extractions in Swedish, both written and spoken. Some examples from this article are given in (1). We mark the relative clause and the gap site of the extracted element.

(1) a. *det*$_1$ finns det ingen [som kan hjälpa mig med _$_1$]
it exist EXPL *nobody that can help me with*
'There is nobody who can help me with it.' (spoken)

b. *det*$_1$ var det ingen [som ville _$_1$]
it was EXPL *nobody that wanted*
'There was nobody who wanted to.' (spoken)

c. *där*$_1$ har jag en moster [som bor_$_1$]
there have I an aunt who lives
'I have an aunt who lives there.' (spoken)

d. *tapeterna*$_1$ var det Sven [som valde _$_1$]
*wall paper.*DEF *was* EXPL *Sven that chose*
'It was Sven who chose the wall paper.' (spoken)

e. ... ett oromantiskt namn$_1$ som jag då inte kände någon [som hette _$_1$]
an unromantic name that I then not knew anyone who was called
'*an unromantic name which I did not know anyone at the time who was called.' (novel 1996)

In (1a,b) the pronoun *det* 'it' has been preposed from a relative clause which modifies the indefinite pivot in an existential sentence; we refer to this construction as a *presentational relative*, see Lambrecht (1988) and Chaves and Putnam (2020, p. 27).[3] In (1c) a short deictic adverb *där* 'there' has been preposed from a *have*-construction which is very similar to presentational sentences (Keenan 1987; McCawley 1981). (1d) is a cleft construction where the object *tapeterna* 'the wall paper' has been preposed from a cleft clause; the clefted constituent *Sven* is definite. (1e) is a relative clause where the relativized item *namn* 'name' has been extracted from a relative clause which modifies the object of the lexical verb *känna* 'be acquainted with'.

These types of constructions, preposing out of presentational relatives or cleft constructions and preposing or relativization from an embedded relative, were most common in

Engdahl's (1997) sample.[4] There were a few examples involving preposing of a *wh*-phrase; they all had the form in (2) which resembles a cleft.

(2) vem$_1$ var det ingen [som kände _$_1$?]
who *was* EXPL *nobody* *that* *knew*
'Who did nobody know?' (spoken)

This is not surprising given that *wh*-questions in Swedish often are clefted with the clefted constituent preposed, as in (3) (Brandtler 2019).

(3) vem$_1$ var det _$_1$ [som kom?]
who *was* EXPL *that* *came*
'Who came?'

In (2), the clefted constituent remains in situ and *vem* is linked to a position inside the cleft clause.[5]

The relative rareness of spontaneous examples with preposed *wh*-phrases fits well with the experimental findings in Kush et al. (2018); Kush et al. (2019) and the corpus studies in Müller and Eggers (2022).

The majority of Engdahl's examples involved extractions out of relative clauses where the subject has been relativized, as in the examples shown so far. There were also a few cleft examples with object relatives, as in (4).

(4) matte$_1$ var det bara pappa$_2$ [jag kunde fråga _$_2$ om _$_1$]
maths *was* EXPL *only* *dad* *I* *could* *ask* *about*
'It was only dad that I could ask about maths.'

In some of the examples the preposed phrase was a lexical DP, as in (1d) and (4), but a large proportion of Engdahl's examples involved either an anaphoric pronoun, as in (1a,b), or a light adverb, as in (1c). Engdahl showed that preposing of unstressed pronouns in Swedish is used as a way of establishing a coherent discourse and suggested that this might be what distinguishes the mainland Scandinavian languages from other languages, essentially the hypothesis that we are testing in this article. In the next section we look first at local preposing in Swedish and then return to extractions from relative clauses.

3. Swedish

The type of preposing of pronouns that we have seen in (1a,b,c) is of course not only found in extraction sentences. Local preposing, in the same clause, of object pronouns is quite common in spoken and written Swedish, as well as in Danish and Norwegian. In this section, we present data from Swedish but similar data are found in Danish and Norwegian, see Engdahl and Lindahl (2014).

3.1. Pronoun Preposing

Starting an utterance with a pronoun is very common in spoken language and Swedish is no exception to this. In most cases the pronoun functions as the subject of the sentence but in the mainland Scandinavian languages it is quite common to start with an object pronoun, or a subject pronoun from an embedded clause.[6] In order to find out when this word order is used, we have conducted several studies using the *Nordic Dialect Corpus (NDC)* (Johannessen et al. 2009) which contains recordings and aligned transcriptions of informal conversations in all the Nordic languages (Lindahl and Engdahl Forthcoming).[7]

The Swedish part of the *NDC* consists of 361 184 words, produced by 133 speakers in 37 locations. Consider the following examples, taken from the *NDC*. We identify the interviewer by *int* and the speakers by *s1*, *s2*, etc. The antecedent of the pronoun is underlined and the preposed pronoun is italicized. (.) marks a short pause, and = that the speaker continues on the next line.

(5) a. *int:* men eh hann du gå i skolan nånting då?
but eh had time you go in school.DEF *anything* PART
'However, did you have time to go to school at all?'

b. *s1:* jo *det*$_1$ fick man ju göra _1
yes it got one PART *do*
'Yes, one had to, of course'

(6) a. *s1:* dessa två de hade (.) de hade slagits (.) där nere=
these two they had they had fought down there
'These two, they had been fighting down there'

b. *s1:* och *det*$_1$ tyckte vi [_1 var väldigt spännande]
and it thought we was very exciting
'and we thought it was very exciting.'

In both (5) and (6), the preposed element is the third person singular neuter pronoun *det* 'it'. In (5) the pronoun acts as a *VP anaphor* which refers to the action *gå i skolan* 'go to school', expressed as a VP in the interviewer's question.[8] In (6) *det* is a propositional anaphor, referring to the event just described by the sentence *de hade slagits* 'they had been fighting'. Here *det* has been preposed from an embedded clause. In both utterances, *det* was unstressed.

Preposing of VP and propositional anaphors, as in (5) and (6), is very common in the NDC but there are also examples with a preposed pronoun which refers to a recently mentioned entity, as in (7) (somewhat abbreviated).

(7) *int:* när köpte du din första bil?
'When did you buy your first car?'

a. *s1:* *den*$_1$ köpte jag _1 1980
it bought I 1980
'I bought it in 1980.'

Here the pronoun *den* agrees in gender and number with the antecedent *bil*, which is the only likely referent. The pronoun is unstressed. If it had been stressed, it would have been understood as a demonstrative, *that*, but this interpretation would not have been plausible in this context. Unstressed *den* and *det* are interpreted as personal pronouns whereas the stressed versions function as distal demonstrative pronouns (Faarlund 2019, p. 27).[9] We gloss the unstressed occurrences as 'it' and the stressed ones as 'that'. As we will show in Section 6.1, this type of non-contrastive preposing is hardly used in English.

In (Lindahl and Engdahl Forthcoming) we investigated how preposing of pronouns is used in the Swedish part of the *NDC*. We found that in a large majority of cases the preposed pronoun referred back to something that was newly introduced, an action as in (5), an event as in (6), or an entity, as in (8a). Following Erteschik-Shir (2007) we refer to this pattern as *focus chaining* since the antecedent is (part of) the relational focus of the preceding utterance, i.e., the new information that is asserted or questioned about the topic (Gundel and Fretheim 2004, p. 177). There were also cases where the antecedent of the preposed pronoun was the topic of the preceding utterance, often realized as a subject. This kind of *topic chaining* was mainly found when the antecedent was an entity, as in (8b) (we just provide the English version of the interviewer's question).

(8) *int:* Are you in touch with anyone who did their military service with you?

s1: ja (.) det var två stycken andra plutonsjukvårdare=
yes there were two other paramedics

a. *dom* var från Fagersta=
they were from Fagersta

b. så *dom*$_1$ har jag ganska bra kontakt med _1
so them have I pretty good contact with

'Yes, there were two other paramedics. They were from Fagersta, so I have pretty good contacts with them.'

The first occurrence of *dom* 'they' in (8a) is an instance of focus chaining. Once *dom* is realized as a subject, it becomes the topic of that utterance. Consequently the second *dom* in (8b) is in a topic chain.

In the examples shown above the preposed pronouns are unstressed. There are also examples where an initial pronoun is stressed and receives a contrastive interpretation, as in (9).

(9) *int:* jo sen finns det kontaktlinser också
yes then are EXPL *contact lenses also*
'Yes, then there are contact lenses as well.'

s1: ja nä *DET*$_1$ vill jag inte ha _1 vet du
yes no that want I not have know you

'No, THAT I do not want, you know.'

By stressing the pronoun and negating the utterance, the speaker emphasizes that he definitely does not want contact lenses.

We found that preposing of pronouns is primarily used in contexts where the antecedent and the pronoun formed a focus chain or a topic chain. In 85% of the investigated cases, the pronoun was locally preposed and in 15% we had long preposing from a subordinate clause, as in (6). One effect of the preposing is that the initial pronoun is interpreted as the *aboutness topic* of the utterance (Reinhart 1981) which we can show using Reinhart's rewriting test. We here apply the test to *s1's* utterance in (8b).

(10) a. *s1* said about *them* (the two paramedics) that he had pretty good contacts with them.

The paraphrase (10) works well and we have a good indication that the preposed *dom* in (8b) acts as the aboutness topic for that utterance.

When it comes to relative frequency of object preposing in spoken Swedish, Jörgensen (1976) showed that this varies with the type of spoken interaction. Based on the material in *Talbanken*[10] he found 14% object initial sentences in interviews, and 9% in conversations and debates (Jörgensen 1976, p. 103). In a more recent study the second author investigated the word order patterns in the Swedish part of the *NDC*, Lindahl (under review). To obtain an unbiased sample, 1000 sentences were extracted from the corpus with the only criterion being that the sentence should have a finite verb. The sentences were manually investigated and all declarative main clauses, 712 in total, were further analyzed. In total, 46 of the sentences in that sample, or 6.5%, had a preposed object and of these, 36 were pronominal (see Table 1 in Section 4.1).

Studies of spoken Swedish show that starting a sentence with *det* is the most common pattern and in most cases it functions as an expletive subject (Allwood 1999; Engdahl 2012). In Engdahl and Lindahl (2014) where we specifically looked for preposed object pronouns (including subject pronouns from embedded clauses), 95% of the hits were *det*. The reason for this is that preposing of VP anaphors or propositional anaphors is by far more common than preposing of pronouns with entity antecedents. Mikkelsen (2015) claims that preposing of VP anaphors is actually obligatory in Danish whereas we argue that there is a strong preference for preposing in Swedish but not a grammatical constraint, Lindahl and Engdahl (Forthcoming). According to Bentzen and Anderssen (2019), preposing of VP and propositional anaphors is always an option in Norwegian.[11]

3.2. Extraction from Relative Clauses

As already mentioned, the first author's investigation was limited to 30 authentic examples (Engdahl 1997). For her dissertation, the second author carried out a larger empirical study of extraction from relative clauses in Swedish based on a collection of 270 spontaneous examples, gathered between 2011 and 2016 (Lindahl 2017b). The collection contained 101 spoken examples from everyday conversations, 60 from radio and television, and 109 written examples.[12]

In a representative subset of the spoken preposing examples (Sample B), slightly over half of the examples, 56%, involved preposed pronouns. Just as in the examples with local preposing, the pronouns extracted from relative clauses typically refer back to something which has just been uttered, either by the same speaker or by an interlocutor. The extracted pronouns are thus part of a focus or a topic chain which connects the utterance to the preceding context. A few examples are given in (11)–(13). For space reasons, the preceding context is sometimes given only in English.

(11) men ingen av dom är ju varmblodiga (.) det_1 finns det inga insekter [som är $\__1$]
but none of them are PRT *warm-blooded it is* EXPL *no insects that are*
'However, none of them are warm-blooded, there are no insects that are.' (Lindahl 2017b, p. 1)

(12) *s1:* The text was rather small.
s2: ja men det_1 var det ingen [som klagade på $\__1$] ser du
yes but it was EXPL *no one that complained about see you*
'Yes, but no one complained about it, you know.' (Lindahl 2017b, p. 77)

(13) den allra största delen av befolkningen, bönderna, den_1 var det adeln [som hade
*the biggest part of population.*DEF, *farmers.*DEF, *it was* EXPL *nobility.*DEF *that had*
domsrätt över $\__1$]
jurisdiction over
'It was the nobility that had the jurisdiction over the largest part of the population, the farmers.' (Radio Sweden, 2015) (Lindahl 2017b, p. 91)

These examples resemble the ones shown in Section 2 from Engdahl (1997). Regarding (11) and (12), these are examples of presentational relatives, cf. (1abc), and (13) is a cleft construction, cf. (1d). This example also illustrates a common use of pronoun preposing in Swedish, namely in *left dislocation* where the initial topic is resumed by a preposed agreeing pronoun, see Holmberg (2020) and Lindahl and Engdahl (Forthcoming) for discussion.

The next two examples involve the lexical main verbs *veta* 'know' and *störa sig på* 'be annoyed at'. In (14), the speaker is talking about driving across the US.

(14) ja det är häftigt det_1 vet jag en [som har gjort $\__1$] (Sw.)
yes it is cool it know I one who has done
'Yes, that is cool! I know someone who has done that!' (Lindahl 2017b, p. 126)

(15) alcoholism is not a disease however
det_1 stör jag mej på folk [som säger $\__1$]
it annoy I me on people that say
'People who say that annoy me.' (Lindahl 2017b, p. 89)

In Lindahl's Sample B, 74% were presentational relatives, 8% were clefts such as (13) and 13% were constructions with lexical verbs.[13] In many of the presentational relatives, the relative clauses are short and contain just a finite auxiliary verb, as in (11), or the support verb *göra* 'do', as in (14). These short relative clauses help to identify or restrict the head of the relative clause. Together with the matrix, the head DP and the relative clause must form a coherent comment on the extracted item which is *relevant* in the sense of Grice (1975).[14]

We end this overview of extraction from relative clauses by noting that Swedish speakers have been using such constructions for a long time. In a guide to *Proper Swedish*, Wellander (1939) discusses the fact that Swedish speakers say things such as (16).[15]

(16) *Det*$_1$ hade jag aldrig träffat någon [som hade gjort _$_1$].
that had I never met someone that had done
'I had never met anyone who had done that.' (Lindahl 2017b, p. 28)

3.3. Summary

In this section, we have shown that the use conditions for extraction of pronouns from relative clauses are the same as for local and non-local pronoun preposing in Swedish.[16] The pronouns are typically part of focus or topic chains; they refer to an event, a property or an entity that has either just been introduced or is already established as a topic. This resembles a well-known strategy for cohesion in discourse, namely to start with a subject pronoun which is part of an anaphoric chain with some element in the preceding utterance (Daneš 1974; Erteschik-Shir 2007). What seems to be special about speakers of Swedish and the other mainland Scandinavian languages is that they use this strategy also for non-subject pronouns, as becomes clear when we look at spontaneous conversations. By preposing a pronoun in a focus or topic chain, the speaker ensures that it will be understood as the aboutness topic for the upcoming utterance. This may in turn explain why preposing the VP or propositional anaphor *det*, as in (5) and (6), is especially common. By doing so the speaker signals that *det*'s antecedent is what the sentence is about.

We noted earlier that preposed pronouns in Swedish are often unstressed; when a pronoun is stressed, it often invokes contrast and the presence of alternatives, see (9). Lexical DPs can also be preposed or extracted (44% in Lindahl's Sample B). These always carry stress and normally invoke alternatives.[17]

4. Icelandic

When Icelandic is mentioned in connection with extractions from relative clauses, it is mainly to establish that such extractions are ungrammatical (Thráinsson 2007; Zaenen 1985). There are no reports of spontaneous extractions in Icelandic in the literature. According to our hypothesis, this might indicate that preposing is not used as a method for establishing cohesion in the discourse. We start by investigating local pronoun preposing in spoken Icelandic and then turn to extractions from relative clauses.

4.1. Pronoun Preposing

Observations of informal conversations between Icelanders give the impression that they use preposing of pronouns much more seldom than Danes, Norwegians or Swedes. In order to investigate whether this impression is correct, we looked at the Icelandic part of the *NDC* (94 338 token, 48 speakers). We found a few examples of object preposing, all of which involve the pronoun *það* 'it, that', as shown in (17) and (18).[18]

(17) a. *s1:* maður fæ- getur ekki notað kreditkort alls staðar
one ge- can not used credit card all places
'One cannot use credit cards everywhere.'

b. *s2:* *það*$_1$ gerum við _$_1$ *það*$_1$ gerum við strákarnir _$_1$ sko
that do we that do we boys.PL.DEF.NOM PRT
'We do, me and the boys do, you know.'[19]

(18) a. *s1:* reyndar býr Clinton í Harlem hverfinu vissir þú það?
actually lives Clinton in Harlem block.DEF.DAT *knew you that*
'Clinton actually lives in Harlem. Did you know that?'

b. *s2:* *það*$_1$ vissi ég ekki _$_1$
that knew I not
'I didn't know that.'

In (17) the preposed *það* is a VP anaphor and in (18) *það* refers back to a proposition, the new information just conveyed.[20] They resemble the Swedish examples we saw in (5) and (6) and are both examples of focus chaining.

This type of preposing of *það* can also be found in written Icelandic and is judged to be natural by many speakers in the acceptability study reported in Lindahl (2022). However, there are no examples of preposed pronouns with entity antecedents in the *NDC* and when asked about such examples, Icelanders tend to supply a contrastive context, see (19) from Engdahl and Lindahl (2014), provided by Halldór Ármann Sigurðsson.

(19) *A:* Have you seen Olaf?

B: nei *hann*$_1$ hef ég ekki séð $\__1$ í allan dag
no him have I not seen in all day

en ég sá konuna hans núna rétt aðan
but I saw wife his now right before

svo að hann hlýtur að vera hérna einhvers staðar
so that he must to be here some where

'No, I have not seen him all day, but I saw his wife just now so he must be somewhere around here.'

The referent *Olaf* is newly introduced in A's question so this is also a case of focus chaining. Note that B contrasts *hann* 'him' with *konuna hans* 'his wife'. This suggests that preposing of entity pronouns is mainly used when the referent is contrasted with some other element.[21]

The impression that object preposing is not very common in Icelandic is confirmed in Lindahl (under review). Using the same method as described for Swedish in Section 3.1, Lindahl extracted 1000 utterances with a finite verb in the Icelandic part of the *NDC* and then investigated all declarative main clauses in the sample.[22] Lindahl categorized the initial constituents in this sample and found substantial differences with respect to the comparable Swedish sample. These are summarized in Table 1.

Table 1. Grammatical function of phrases in the prefield in NDC, from Lindahl (under review).

	Swedish		Icelandic	
	No. of Instances	**Percent**	**No. of Instances**	**Percent**
Subject/expletive	421	59.1%	400	75.2%
Adverbial	177	24.0%	87	16.4%
V1 declarative	32	4.5%	37	7.0%
Object	46	6.5%	1?	0.2%
Other	36	5.8%	7	1.3%
Total	712	100%	532	100%

First we note that the proportion of subject initial utterances was much larger in Icelandic, 75.2% compared to 59.1%, whereas preposing of adverbials was more common in Swedish. Verb initial (V1) declaratives, was somewhat more common in Icelandic than in Swedish. As for preposing of objects, there were actually no unambiguous examples with a preposed object in this random sample of Icelandic utterances, compared to 46 in the Swedish sample. We interpret this as an indication that object preposing is not a particularly common strategy in spoken Icelandic.

4.2. Extraction from Relative Clauses

As mentioned in the Introduction, extraction from relative clauses in Icelandic is considered to be ungrammatical (Thráinsson 2007; Zaenen 1985). However, since these articles only mention examples with extracted lexical DPs, an interesting question is whether Icelandic speakers accept the types of extractions that are found in Swedish (see Section 3.2) even if they do not produce them spontaneously. The second author reports on such a study in Lindahl (2022). In two parallel acceptability studies, speakers of Swedish and Icelandic were asked to rate sentences with extractions, using examples similar to the spontaneously produced ones from mainland Scandinavian we have seen in this article. The extracted phrases used in all of the test sentences were *það* in Icelandic and *det* in Swedish, and they all had sentential or VP antecedents. The participants were asked to rate examples such as (20), using the scale natural, somewhat strange and unnatural.

(20) Þú getur notað gjafakortið til að kaupa bíómiða og *það*$_1$ eru margir [sem gera _$_1$].
*you can use voucher.*DEF *to to buy movie ticket and that are many who do*
'You can use the voucher to buy a movie ticket, and there are many people who do.'

In brief, Lindahl found that extraction from relative clauses was rated very poorly in the Icelandic part of the study, and clearly much worse than the parallel Swedish examples. Local preposing was judged to be natural by two thirds of the participants. Long preposing from an *að*-clause received mixed ratings, but was rated worse than good fillers. In Swedish, preposing from *att*-clauses was rated on a par with good fillers. As for extraction from relative clauses, the Icelandic speakers found extraction unnatural in all of the test sentences, regardless of the type of relative clause. The Swedish speakers rated test items with a presentational relative clause as natural sounding more often than those items which contained a lexical verb. The study thus confirms previous reports that extraction from relative clauses is not acceptable in Icelandic.

4.3. Summary

Preposing of the VP or propositional anaphor *það* is sometimes used in Icelandic, primarily in local clauses, in a way that resembles the Swedish pattern, albeit much less frequently. However, when pronouns with entity antecedents are preposed, they reportedly receive a contrastive interpretation, just as in English (see Section 6.1). The low occurrence of preposed pronominal objects in Icelandic suggests that preposing is not used as a natural way of connecting utterances in the same way as in mainland Scandinavian. Since extraction from relative clauses seems to be unacceptable, we think this supports our hypothesis that there is a connection between using pronoun preposing as a way of connecting utterances and a willingness to accept long extractions. The study reported in Lindahl (2022) confirms that there is indeed a difference between Icelandic and Swedish in how long preposing of anaphoric pronouns is judged—such preposings are rated as natural more often in Swedish than in Icelandic, in line with our hypothesis. In addition, Lindahl found a clear effect of clause type in Icelandic. While a substantial part of the Icelandic participants rated long preposing from *að*-clauses as natural, even though such preposings are not common in spoken Icelandic, they did not do so for extraction from relative clauses. This suggests that there is also a structural constraint which blocks extraction from relative clauses in Icelandic.

5. Faroese

Turning to Faroese, very little is known about possible uses of pronoun preposing in Faroese and extraction from relative clauses has not been discussed much in previous research, apart from when it is pointed out that it is unacceptable (Platzack 2014; Thráinsson et al. 2004).

5.1. Pronoun Preposing

In the Faroese part of the *NDC* (64803 tokens from 20 speakers) there are more than a hundred examples with a preposed *tað* 'it, that', pronounced /tæ/, including the two following examples.

(21) a. *s1:* hevur tú spælt gekk nú tá ið gekkaupphæddin var so stór
have you played Gekk now then REL payout sum was so large

b. *s2:* ja (.) veitst tú hvat $tað_1$ havi eg faktiskt _1
yes know you what that have I actually

'Do you know what, I actually have.'

(22) altso tað er ordiliga hugnaligt $tað_1$ haldi eg _1
PRT it is really nice it think I

'It's really nice, I think so.'

In (21) *tað* is a VP anaphor and in (22) *tað* refers back to a proposition. There are notably fewer preposed pronouns with entity antecedents, as in (23).

(23) Carl Johan Jensen hann eg veit at hann er rithøvundur (.)
Carl Johan Jensen he I know that he is author

$hann_1$ veit eg einki um _1
him know I nothing about

'Carl Johan Jensen, I know he is an author. I do not know anything about him.'

Since there is so little data available on spoken Faroese, the second author carried out a controlled production study in Faroese.[23] The study took advantage of the fact that practically all Faroese speakers know Danish well; Danish is taught in schools from third grade and used frequently in the society. In the study, 91 native Faroese speakers were asked to translate Danish sentences into Faroese. 82 of the participants were high school students, and 9 were between 38 and 69 years at the time. The Danish sentences were chosen so as to resemble spoken dialogue.

An example of the task is shown in (24). Each sentence was given in a context, also in Danish, which the participants could choose to translate if they wanted to. In the examples below, we have underlined the antecedent and italicized the pronoun, but this was not performed in the experiment.

(24) Anna spurgte os, hvad klokken var, men det_1 vidste vi ikke _1. (Da.)
*Anna asked us what clock.*DEF *was but it knew we not*

'Anna asked us what time it was but we did not know.'

The Faroese speakers had no difficulties in translating this example. Their answers varied, as shown in (25). The percentages to the right indicate how often the participants used this word order in their translations.[24]

(25) Anna spurdi okkum hvat klokkan var ... (Fa.)
*Anna asked us what clock.*DEF *was*

a. men $tað_1$ vistu vit ikki _1 82%
but that knew we not

b. men vit vistu *tað* ikki 2%
but we knew that not

c. men vit vistu ikki Ø 14%
but we knew not

In this example, the preposed *det* in the Danish version is a propositional anaphor. The Faroese participants all translated it using *tað* and in 82% of the answers they preposed it. The same tendency showed up in examples where a VP anaphor was preposed from an embedded *at*-clause. The Danish original is shown in (26) and in (27) we give the relevant parts of the Faroese translations of the example.

(26) Hvis jeg ikke tager fejl, og *det*$_1$ tror jeg ikke, at jeg gør _$_1$, har Ole satt kagen
*if I not take error and that think I not that I do has Ole put cake.*DEF
i køleskabet. (Da.)
*in fridge.*DEF
'If I'm not mistaken, and I do not think I am, Ole has put the cake in the fridge.'

(27) a. vissi eg ikki takið fel, og *tað*$_1$ haldi eg ikki at eg geri _$_1$ 67% (Fa.)
if I not take error and that think I not that I do

b. um eg ikki taki feil$_1$, sum eg ikki haldi eg geri _$_1$ 13%
if I not take error which I not think I do

c. vissi eg ekki taki feil, och *tað*$_1$ haldi eg ikki _$_1$ 8%
if I not take error and that think I not

d. um eg ikki taki feil$_1$, sum eg ikki geri _$_1$ 4%
if I not take error which I not do

e. um eg ikki taki feil, og eg haldi ikki at eg geri Ø 8%
if I not take error and I think not that I do

In total, 92% of the Faroese participants translated the Danish sentence using a preposing strategy, either a local or long preposing of *tað* or relativization with *sum*. However, when the Danish sentence contained a preposed entity pronoun as in (28), the Faroese participants preferred a different strategy.

(28) Jeg var hjemme hos min bror i går. *Han* bor tæt på mig,
I was home my brother yesterday he lives close at me
så *ham*$_1$ ser jeg tit _$_1$. (Da.)
so him see I often

'I was at my brother's yesterday. He lives close to me, so I see them often.'

(29) Eg var heima hjá beiggja mínum í gjar. *Hann* býr tætt við hjá mær ... (Fa.)
I was home brother mine yesterday he lives close at by me

a. so *hann*$_1$ síggi eg ofta _$_1$ 20%
so him see I often

b. so eg síggi *hann* ofta 59%
so I see him often

c. so vit síggjast ofta 19%
*so we see.*RECIP *often*

For this example, only 20% of the Faroese translations replicate the preposing from the Danish original. Instead, most participants chose to start with the subject as in (29b,c). In total, 19% of the translations used an alternative formulation with a reciprocal form of the verb, as in (29c). The Faroese speakers' preference for avoiding preposing was even more pronounced when the Danish original involved long preposing of an entity pronoun from an *at*-clause.

(30) Tove: Hvor er kagen? (Da.)
*where is cake.*DEF

Mette: *Den*$_1$ tror jeg, at Ole satte _$_1$ i køleskabet.
*it think I that Ole put in fridge.*DEF

'I think Ole put it in the fridge.'

(31) Hvar er kakan? (Fa.)
*where is cake.*FEM.DEF
'Where is the cake?'

a. *Hana*$_1$ haldi eg, at Óli setti _$_1$ í køliskápi 4%
her think I that Óli put in fridge.DEF

b. Eg haldi at Óli setti *hana* í køliskápið 79%
I think that Óli put her in fridge.DEF

c. Eg haldi at Óli setti Ø í køliskápi 17%
I think that Óli put in fridge.DEF

Only 4% of the translations replicated the preposing of an entity pronoun. The Faroese participants clearly preferred to leave the pronoun in the verb phrase or leave it out. Given the different strategies used by the Faroese participants, it seems that preposing of the VP or propositional anaphor *tað* is a common strategy in Faroese, but it does not extend to entity pronouns.

There are both advantages and disadvantages with using a translation method. It can be a useful method for finding examples that are hard to locate in corpora and it can elicit natural sounding Faroese exchanges, especially if the materials to be translated come from natural dialogue. One disadvantage is that participants may transfer some aspects of the Danish originals into Faroese. However, when the Faroese translations deviate from the Danish originals, this is most likely an indication that the participants do not recognize this as an acceptable word order in Faroese. For example, only a few of the translations retained the preposed word order when the pronoun referred to an entity, see (29a) and (31a).

As for the frequency of object preposing, we can get at least an idea about it from a smaller sample of 300 sentences from the Faroese part of the *NDC*. The sample contained 192 declarative main clauses, and out of these, 15 started with a preposed object, in 12 cases *tað*, which was the only pronoun found. Although the material is smaller, the rate of object preposing, 7.8%, is comparable to what Lindahl found using the same type of search in Swedish, 6.5%, see Table 1. Object preposing thus seems to be fairly common in spoken Faroese and definitely more common than in Icelandic.

5.2. Extraction from Relative Clauses

According to the literature, extraction from relative clauses is not acceptable in Faroese. Platzack (2014) gives the following example, which is a translation of a Swedish example in Allwood (1976).

(32) * *Slíkar blómur*$_1$ kenni eg ein mann [sum selur _$_1$] (Fa)
such flowers know I a man who sells

(Platzack 2014, p. 10)

However, we have noticed a handful examples in Faroese newspapers, two of which are shown here.

(33) Tað ber eisini til at koma til viðgerð saman við fyrrverandi makanum, og
it bears also to to come to treatment together with former spouse.SG.DEF *and*
tað$_1$ eru tað nógv [sum gera _$_1$] (Fa.)
that are there many who do
'It is also possible to undergo treatment together with your former spouse, and there are many people who do.' (*Dimmalætting*, 24 April 2015)

(34) *"Kjakokrati"*$_1$ var tað onkur, [sum rópti tað _$_1$], tá fólk á ymsum
debate-ocracy was there no one who called it when people on various
internetsíðum viðmerkja evnir, ella geva sína meining til kennar. (Fa.)
web pages comment topics, or give REFL *meaning to known*
'No one called it "debateocracy" when people on various web pages commented on topics or let their options be known.' (*Dimmalætting*, 1 March 2019)

In order to investigate whether such constructions are used productively, Lindahl included extractions from relative clauses in her translation study. The 91 Faroese speakers were

asked to translate five examples with relative clauses. Three of them were similar to (35) where the relative clause modifies the object of a lexical verb, here *ved* 'know'.

(35) Jeg synes sommeren er den bedste tid på året, og det_1 ved jeg mange,
*I think summer.*DEF *is the best time on year.*DEF *and it know I many*
[som er enige med mig om _1]. (Da.)
that are in agreement with me about
'I think that summer is the best time of year, and I know many people who agree with me about that.'

(36) Eg haldi at summarið er besta tíðin á árinum, (Fa.)
*I think that summer.*DEF *is best time.*DEF *on year*

a. og $tað_1$ veit eg nógv [sum er einigur við meg um _1] [1 inf.]
and that know I many that are in agreement with me about

b. og $tað_1$ veit eg, at tað eru nógv, [sum eru samd við mær í _1][1 inf.]
and that know I that there are many that are same with me in

c. og $tað_1$ veit eg at nógv eru samd um _1
and that know I that many are in agreement about

d. og eg veit at fleiri eru samd við mær
and I know that many are in agreement with me

e. og eg veit fleiri, sum er samd við mær
and I know many who are in agreement with me

f. og eg veit nógv ið eru einig við mær
and I know many who are in agreement with me

g. og eg veit nógv ið eru einig við mær í *tí*
*and I know many who are in agreement with me in that.*DAT

Only two of the Faroese participants (2%) replicated the preposing in the Danish original and produced the versions in (36a,b). The rest of them used a variety of strategies to convey the content of the Danish sentence.[25] In (36c), *tað* is fronted, but the sentence has been reformulated using an *at*-clause instead of a relative clause. (36d,e,f) use slight reformulations leaving out the pronoun, and (36g) keeps the relative clause, but leaves the pronoun *tí* (the dative form of *tað*) in situ.

For the two sentences where the Danish original was a presentational relative, as in (37), the Faroese speakers did produce some extractions from relative clauses in their translations, see (38).

(37) Ole undrede sig over, om det ville regne i morgen, men det_1 var der ingen,
Ole wondered REFL *over if it would rain tomorrow but that was there no one*
[der troede _1]. (Da.)
who believed
'Ole wondered whether it would rain tomorrow, but no one thought so.'

(38) Óli ivaðist í, um tað fer at regna í morgin ... (Fa.)
Óli doubted in whether EXPL *go to rain tomorrow*

a. men $tí_1$ var tað eingin [ið trúði _1] [1 inf.]
*but that.*DAT *was there no one who believed*

b. men $tað_1$ var tað ongin [sum helt _1] [16 inf.]
but that was there no one who thought

c. men tað var ongin [sum helt *tað*]
but there was no one who though that

d. men $tað_1$ helt ongin _1
but that thought no one

e. men *taðˌ*$_1$ trúði ongin _$_1$
but that believed no one

f. men *taðˌ*$_1$ roknaði ongin við _$_1$
but that reconned no one by

g. men ongin helt *tað*
but no one though that

h. men tað var eingin ið trúði honum
but there was no one who believed him

A total of 17 participants (19% of the replies) retained the Danish structure, as shown in (38a,b). One participant used the relativizer *ið* as in (38a) and 16 participants used the relativizer *sum*. The other 74 informants reformulated the sentence either by not using a relative clause in the translation, by leaving the object in situ, as shown in (38c–g), or both, or using other reformulations, as in (38h). In the two examples where the matrix verb was *vera* 'be', 10–20% of the translations retained the Danish extraction structure, compared to 2–4% when the verb was *vita* 'know' or *kende* 'be acquainted with'.[26] This suggests that this particular type of extraction sentence is less degraded in Faroese, and may even be acceptable for some speakers, while extraction from other relative clauses is unacceptable. Faroese could thus potentially be a language that shows evidence for the "pseudo-relative" hypothesis put forth by McCawley (1981). He proposed that the relative clause in existential sentences in English is not a true relative clause; instead it has a somewhat reduced structure. This could explain why extraction in such an environment is sometimes marginally acceptable in English (see also Chung and McCloskey 1983; Kush et al. 2013).

5.3. Summary

There are clearly similarities between Faroese and Icelandic. In both languages the VP or propositional anaphor *það,tað* can be preposed locally and from *að/at*-clauses, as shown in the acceptability and translations studies summarized above. However, the impression is that this is more common in Faroese. Preposing of entity pronouns seems not to be a natural strategy given the alternative translations provided in (29) and (31). Given that preposing of *tað* is fairly common in Faroese, one might expect speakers to use preposing of *tað* from relative clauses as well. The translation study showed that this is not the case, except for preposing in presentational relative clauses, as in (38). This is also the type of preposing that Faroese speakers are most likely to hear and read since this is the most common type in Danish (Müller and Eggers 2022). In this respect, the Faroese speakers differ from Icelandic speakers who judge extraction from presentational clauses to be unnatural to the same extent as extractions from other types of relative clauses, as discussed in Section 4.2.

6. Comparisons with Other Languages

We have suggested that the frequent use of preposing of unstressed anaphoric pronouns in Swedish is a crucial factor for explaining the presence of spontaneous extractions from relative clauses in this language. Since extraction from relative clauses is not found in English and the continental Germanic languages German and Dutch, it becomes relevant to investigate whether local and long preposing is used in these languages.

6.1. English

In English, preposed (topicalized) phrases are normally stressed and are understood to imply contrast. Exceptions to this are the light adverbs *then* and *there* which do not require stress in initial position when they serve to connect the utterance to the preceding context. Attempts to prepose an unstressed personal pronoun sound very strange. Compare the options for answering the question in (7) in English.

(39) When did you buy your first car?
a. **It*$_1$ I bought _$_1$ in 1980.

b. I bought *it* in 1980.

c. #THAT$_1$ I bought _$_1$ in 1980.

The preposed version in (39a) is clearly unacceptable whereas the version with the pronoun in situ in (39b) is fine. Preposing a stressed demonstrative is grammatical, (39c), but hardly appropriate in this context.[27]

When we look at the Swedish examples involving the VP anaphor *det*, we find that they are best rendered in English using VP ellipsis (Ø) (Bentzen et al. 2013; Hankamer and Sag 1976; Hardt 1999). Consider the question in (40).

(40) Have you been to Oslo?

a. Ja, *det*$_1$ har jag _$_1$.
yes it have I

b. Ja, jag har *det*.
yes I have it

c. Yes, I have Ø.

In Swedish, this is a context where the VP anaphor *det* typically would be preposed. Leaving it in situ, (40b), is not ungrammatical but the preposed version is preferred (Lindahl and Engdahl Forthcoming).[28] In English, repeating the finite verb and leaving out the rest of the VP is common and this is also what we have done in the English translations of the examples with VP anaphors, Swedish (5), Icelandic (17) and Faroese (21). Note that the VP ellipsis strategy can also be used in embedded clauses in English where Swedish uses preposing.

(41) Do you think I should go to Oslo?

a. Ja, *det*$_1$ tycker jag du ska _$_1$.
yes it think I you shall

b. Yes, I think you should Ø.

There are clearly similarities between VP ellipsis in English and VP anaphor preposing in Swedish; in particular both require that the antecedent VP is available in the immediate context.

In contexts where a preposed *det* in Swedish refers back to a recently expressed proposition, the English version can sometimes have the anaphor *so*, as in (42) from the *NDC*.

(42) a. *int:* tycker du det är roligt med små barn?
think you it is fun with small children
'Do you think small children are fun?'

b. *sp1:* ja *det*$_1$ tycker jag faktiskt _$_1$
yes it think I actually
'Yes, I actually think *so*.'

Given that VP ellipsis can also be used inside embedded clauses in English, as in (41b), one might ask whether this strategy can be used in English translations of extraction clauses. We have actually done so in (11) and repeat the translation here in (43).

(43) However, none of them are warm-blooded, there are no insects that are Ø.

This long distance VP ellipsis appears to be acceptable at least to some English speakers. However, the rest of the Swedish extraction examples, which involve preposed propositional or entity pronouns, are best translated into English with the pronoun in situ.[29]

6.2. German and Dutch

German and Dutch are both verb second languages and local preposing is quite common, but there are differences compared with Swedish. Consider the German version of the car dialogue.

(44) Wann hast du dein erstes Auto gekauft?

a. *Das*$_1$ habe ich _$_1$ in 1980 gekauft.
that have I in 1980 bought

b. Ich habe *es* in 1980 gekauft.
I have it in 1980 bought

c. **Es*$_1$ habe ich _$_1$ in 1980 gekauft.
that have I in 1980 bought

The unmarked answer would be with a preposed unstressed demonstrative pronoun, a so called *d*-pronoun, as in (44a), or an unstressed personal pronoun in situ, (44b). Preposing of the neuter pronoun *es* is very restricted, see the discussions in Frey (2006); van Craenenbroeck and Haegeman (2007) and Theiler and Bouma (2012). Corpus studies reveal that preposing of personal pronouns is uncommon compared to preposing of *d*-pronouns. In a newspaper corpus, it was much more common for object *d*-pronouns to be preposed than to appear in situ whereas no personal object pronouns were preposed (Bosch et al. 2007).

In Dutch, preposing of personal pronouns is also uncommon compared to preposing of demonstratives (*d*-pronouns) (van Kampen 2007).

(45) Wanneer heb jij je eerste auto gekocht?

a. *Die*$_1$ heb ik _$_1$ in 1980 gekocht.
that have I in 1980 bought

b. Ik heb *hem* in 1980 gekocht.
I have it in 1980 bought

c. #*Hem* heb ik _$_1$ in 1980 gekocht.
it have I in 1980 bought

As in German, the unmarked answers have a preposed *d*-pronoun or a personal pronoun in situ. Preposing an unstressed personal pronoun is not ungrammatical but pragmatically odd.[30] Bouma (2008, p. 112) found 4 preposed object personal pronouns in a 9 million words corpus of spoken Dutch (*Corpus Gesproken Nederlands*), compared to 2723 preposed *d*-pronouns.

When it comes to non-local preposing, both German and Dutch restrict this to contrastively stressed phrases. This has led some German linguists to propose that there are two types of preposing. Werner Frey distinguishes what he calls *formal movement* from *true A-bar movement* (Frey 2004, 2006, 2007). Formal movement involves local preposing to Spec,CP of the highest constituent in the middle field (IP). This accounts for preposing of subjects, both referential and expletive. Objects can also be preposed by formal movement when they are topics. On Frey's analysis this means that they have first moved to a topic position above the subject position in the middle field; this way they become the highest constituent in IP. True A-bar movement accounts for all other movement into Spec,CP and is linked to a marked interpretation, typically involving contrast. A similar proposal is made by Fanselow (2016) who distinguishes *unrestricted V2* where there are no pragmatic effects of the preposing and *restricted V2* which comes with pragmatic effects and induces contrast.

This distinction seems to capture the situation in German. All examples of spontaneously produced long preposings in German that we are aware of are contrastive, as predicted by Frey and Fanselow, see Andersson and Kvam (1984) and Lühr (1988).[31] Two examples from Andersson and Kvam (1984) are shown here, involving preposing of contrastive adverbial phrases from a *daß* 'that' clause and an *ob* 'if' clause.

(46) das könnte vielleicht regional verschieden sein und z.B. in HanNOver$_1$ würde
that could maybe regionally different be and, e.g., in Hannover would
ich zweifeln, daß jemand das _$_1$ sagt
I doubt that someone that says
'That could perhaps differ regionally. For instance, in Hannover, I would doubt that anyone says that.'

(47) also DIENstag$_1$ weiß ich nicht genau, ob er _$_1$ kommt, doch MITTwoch$_1$ ist er
so Tuesday know I not exactly if he comes, but Wednesday is he
ganz bestimmt _$_1$ da
all certain there
'Tuesday, I'm not sure if he will come, but Wednesday, he will certainly be there.'

To some extent, the distinction between formal movement and true A-bar movement is relevant for Swedish as well since expletive subjects can only be preposed in the local clause. However, we find no evidence for a topic position above the subject in the middle field in Swedish which means that preposing of unstressed object pronouns cannot be handled by formal movement (Lindahl and Engdahl Forthcoming). Instead, it seems that A-bar movement in Swedish is not correlated with contrast or a marked interpretation but can be used in order to connect utterances involving focus and topic chains.[32] Highly proficient L2 speakers of German with Swedish L1 sometimes extend this type of preposing into spoken and written German with the result that "the preposed item is understood to be highlighted as very important even when it is not", as a native speaker of German recently commented.[33]

7. Concluding Remarks

In this article, we have explored the hypothesis that frequent use of local preposing of object pronouns increases the likelihood that speakers will also use long preposing from embedded clauses as well as from relative clauses. For Swedish, it seems clear that preposing of both subject and non-subjects pronouns is a common strategy when the pronoun refers to a recently mentioned event, property or entity, that is when they are part of a focus or topic chain. This preposing strategy can be extended to subordinate clauses and to relative clauses, provided that the rest of the utterance is a relevant comment on the preposed item in the context. Acceptable long preposings and extractions are thus doubly context dependent; first, the preposed item must be part of a focus or topic chain with respect to the previous utterance and second, it must be possible to interpret the preposed item as the aboutness topic of the utterance with the rest providing a comment that meets Grice's Maxim of Relevance. The second property presumably lies behind the fact that short presentational relative clauses are most common. When both these pragmatic conditions are met, long extractions can be used and do not cause problems for the listeners/readers.[34] This does not mean that there are no syntactic constraints on extractions in Swedish. For instance, the Coordinate Structure Constraint seems to apply, see, e.g., the discussion in Lindahl (2017b).

The situation in Icelandic is clearly different. Local preposing of the VP or propositional anaphor *það* occurs but is used much more seldom than in Swedish. Preposing of entity pronouns is unusual and seems to have a contrastive effect instead of being used for cohesion. Furthermore the pragmatic conditions that are important in Swedish do not seem to play any role when it comes to extraction from relative clauses. Not even when the sentences were pragmatically very plausible did the Icelandic participants in Lindahl's (2022) study judge them to be natural. We conclude that there must be a structural constraint operating in Icelandic which prevents such extractions and which is not affected by pragmatic conditions.

Faroese presents a more mixed picture. On the one hand, Faroese speakers often prepose the VP or sentential anaphor *tað*, similar to Danish. This suggests that preposing is employed as a cohesive device, but note that it does not seem to extend to entity pronouns.

On the other hand, Faroese speakers react more similar to Icelanders when it comes to extractions from relative clauses; they try to avoid using them when they translate such sentences from Danish. This suggests that extractions from relative clauses are unacceptable in Faroese, similar to Icelandic, presumably due to a structural constraint. However, there is one interesting difference between the Faroese and Icelandic speakers present in our studies. The Faroese speakers seem to accept, and occasionally produce, extractions from presentational relatives, whereas the Icelanders find these unnatural. This may be due to influence from Danish where this kind is the most common type of relative clause extraction (Müller and Eggers 2022). It may also be that these relatives have a reduced structure in Faroese which makes extraction more acceptable. More research on Faroese is clearly called for.

One important aspect of pronoun preposing in the mainland Scandinavian languages is that it often involves unstressed pronouns. If the antecedent of the pronoun has just been introduced or already is a topic in the previous utterance, it is often natural to continue with an unstressed pronoun. (This holds for both subject and non-subject pronouns). In English, object preposing tends to induce a contrastive interpretation and consequently the preposed item must be stressed; this holds both for local and non-local preposing. In Dutch and German, unstressed demonstrative pronouns can be preposed from the same clause but only contrastively stressed items can be preposed from embedded clauses. In English, one way of connecting utterances without invoking contrast is to use VP ellipsis. This often turns out to be the best translation of Swedish utterances with preposed pronouns, but there are limitations, especially with regard to relative clauses.

Our comparative investigation has shown that preposing in some languages can be used to connect the utterance to the preceding context through anaphoric chains. This means that (long) preposing by itself is not necessarily associated with a contrastive interpretation. Whether a contrastive interpretation is plausible depends rather on the context of use and the type of preposed item. This should have consequences for psycholinguistic investigations of extractions in languages such as Swedish. If, as has been common so far, the experimenter uses materials with preposed lexical DPs, then very often a contrastive reading emerges. However, if the context calls for a cohesive continuation, a focus or topic chain, then a sentence with a preposed pronoun might be more natural.

Author Contributions: E.E. is mainly responsible for writing Sections 1, 2, 3.1, 6 and 7. The remaining Sections were written together. F.L. carried out the corpus studies reported in Sections 3–5 as well as the experimental studies in Sections 4–6. All authors have read and agreed to the published version of the manuscript.

Funding: Part of this research was supported by an International Postdoc grant (Dnr. 2017-06139) from the Swedish Research Council, awarded to F.L.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Nordic Dialect Corpus http://www.tekstlab.uio.no/nota/scandiasyn/. The Swedish data concerning extraction from relative clauses are available at https://svn.spraakbanken.gu.se/sb-arkiv/pub/lindahl/2017/. The Icelandic and Faroese data presented in this study are available on request.

Acknowledgments: Earlier versions of parts of this paper have been presented at the workshop on Word Order in the Scandinavian languages, University of Konstanz, December 2018, the Linguistics Research Seminar at the University of Gothenburg, May 2021, and at the 12th International Conference of Nordic and General Linguistics, University of Oslo, June 2021. We would like to thank Maia Andréasson, Miriam Butt, Nomi Erteschik-Shir, Benjamin Lyngfelt, Peter Sells and Annie Zaenen for discussion and comments. We are grateful to Kristin Hagen and Anders Nøklestad from the Text laboratory at the University of Oslo for help with using the *Nordic Dialect Corpus*. We also thank Victoria Absalonsen, Gunnvør Hoydal Brimnes, Zakaris Svabo Hansen, Jógvan í Lón, Jóhannes Gísli Jónsson, Ida Larsson, Solveig Malmsten, Hjalmar Petersen, Sigríður Sigurjónsdóttir, and Annika

Simonsen for help with the Danish, Faroese and Icelandic questionnaires and example sentences and for contacts with test participants. For the final version of the article, the comments from two anonymous reviewers were very helpful.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 See Chaves and Putnam (2020, p. 67) for additional references.

2 We use the term *preposing* rather than the commonly used term *topicalization* in order to distinguish the syntactic positioning from any information structural effects this may have, see Ward (1985) and the discussion in (Lindahl and Engdahl Forthcoming).

3 Swedish data are discussed in Sundman (1980, p. 59) and Teleman et al. (1999, volume 3, p. 423f.).

4 Löwenadler (2015) suggests that these common types should be seen as conventionalized constructions.

5 Søfteland (2013) calls this type a *presentational cleft* since the clefted constituent *ingen* 'nobody' is indefinite, as in presentational sentences, see (1b). She distinguishes them from *it*-clefts where the clefted constituent is normally definite, as in (1d).

6 Swedish is a verb second language and we assume that both subjects and non-subjects are preposed to a Spec position in the C domain in declarative clauses, which we refer to as the *prefield*. In (Lindahl and Engdahl Forthcoming) we develop the syntactic analysis further, adopting the bottleneck hypothesis in Holmberg (2020).

7 The sound files along with transcripts with word by word translations into English can be accessed on the web page of the Text Laboratory at the University of Oslo (http://www.tekstlab.uio.no/nota/scandiasyn/ accessed on 1 March 2022). You can search for the utterances, listen to the sound files and see the examples in context. To facilitate the search, the transcripts use standard orthography which we have retained in the examples cited, but added underlining, italics and gap locations. The Norwegian and Eldfdalian transcripts in addition contain a simplified phonetic transcription.

8 On VP anaphora in Scandinavian, see, e.g., Ørsnes (2011); Lødrup (2012) and Mikkelsen (2015).

9 Personal pronouns include first, second and third person referential pronouns. The pronoun *det* also functions as a non-referential *expletive*, in which case it cannot be stressed.

10 *Talbanken* (96,346 words) was collected in Lund in the 1970s. The materials are available in *Språkbanken* and can be searched using the search engine *Korp* (https://spraakbanken.gu.se/korp, accessed on 1 March 2022.)

11 Mikkelsen (2015); Bentzen and Anderssen (2019) and Lindahl and Engdahl (Forthcoming) also discuss interactions with Object shift.

12 See Lindahl (2017b, p. 45ff.) for details about the data collection.

13 The relative frequency for lexical verbs, 13%, is much higher in Lindahls's spoken collection than what Müller and Eggers (2022) find in their corpus study of written Danish (7/940). We suspect that a similar study of written Swedish would would also find a higher proportion of presentational relatives.

14 Establishing what is a relevant comment in a particular context is difficult, see Allwood (1976); Andersson (1982) and Engdahl (1982) for some attempts. Recently Chaves and Putnam (2020, p. 120) have introduced the term *Relevance Islands* for contexts which are sensitive to Gricean maxims.

15 The earliest mention of this type of extraction that we are aware of is in Mikkelsen (1894), an early grammatical descriptions of Danish. See Lindahl (2017a, p. 27) for some of his examples.

16 The same probably holds for extraction from embedded interrogatives. In Lindahl's (2022) acceptability study, the participants rated sentences where the VP or propositional anaphor *det* had been extracted from *wh*-clauses as natural 65.7% of the time.

17 Lindahl (2017b, p. 146f.) also discusses examples where an extracted stressed item is understood as the focus of the utterance, typically in answers to questions.

18 The propositional or VP anaphor *það* is usually glossed as 'that' in the literature on Icelandic and we follow this in the examples below.

19 See Sigurðsson and Wood (2020) for an analysis of the use of conjoined subjects as in *við strákarnir*.

20 In Object shift contexts, such as (18b), the gap could also be located before the negative adverbial, but nothing hinges on this for our analysis.

21 A reviewer pointed out that examples with preposed object pronouns can be found in the large corpus of written Icelandic, the Gigaword Corpus (https://malheildir.arnastofnun.is/, accessed on 28 April 2022). However, if these sentences had been spoken, the pronoun would have "a regular main-clause initial stress", according to the reviewer. This may be a relevant difference with Swedish. In Lindahl and Engdahl (Forthcoming) we include Praat analyses (Boersma and Weenink 2020) of Swedish examples with preposed unstressed pronouns. These analyses show that the initial pronoun often does not form a separate prosodic phrase but is incorporated into the verb, see Myrberg and Riad (2015). Similar investigations of the prefield in Icelandic are needed as well as more informant studies.

[22] This involved excluding questions, imperatives, embedded clauses and tags. The Icelandic sample contained more embedded clauses and more tags such as *þú veist* 'you know' than the Swedish sample with the result that there were fewer relevant clauses in Icelandic.

[23] A more detailed presentation of this study in Swedish is forthcoming, Lindahl (to appear).

[24] The percentages do not always add up to 100, because there are a few cases where a participant did not contribute a translation. The participants were asked to use the word order they found natural for spoken or informal written Faroese. We are showing the answers exactly how they were written by the participants, including spelling errors and any informal/non-standard spelling. For example, many speakers have chosen to leave out *-ð* in the definite suffix, which is silent in spoken language.

[25] Since we are mainly interested in the extraction cases here, we only give the number of informants who produced such translations.

[26] Note the interesting translation in (36b) where the informant inserts an additional *at*-clause which permits him/her to reformulate it as a presentational relative, thereby avoiding having a relative clause embedded under *vide* as in the Danish original.

[27] (39c) would have been appropriate if the question had involved a narrow focus, *When did you buy your FIRST car?*.

[28] It is possible that the in situ order is used more frequently in Norwegian than in Danish and Swedish, cf. Bentzen and Anderssen (2019). More comparative research is required in order to establish if this is the case and why. There is also an issue whether the VP anaphor can precede negation in *Object shift*, see Mikkelsen (2015); Ørsnes (2013) and Engdahl and Zaenen (2020).

[29] Similarities and differences between movement and ellipsis have been much discussed, see, e.g., Johnson (2001) and Aelbrecht and Haegeman (2012).

[30] The examples and judgments in (45) were supplied by Gerlof Bouma.

[31] The examples in Zifonun et al. (1997) are all taken from Andersson and Kvam (1984). There is considerable discussion concerning long *wh*-movement, especially the so called *was was* construction, see the articles in Lutz et al. (2000).

[32] See Lindahl and Engdahl (Forthcoming) for a detailed discussion of the structure of the C-domain in German and Swedish.

[33] Professor Christiane Andersen, personal communication.

[34] We have looked at extended contexts for the attested examples we have investigated and have not seen any evidence that they are difficult to produce, for the speaker, or to understand, for the addressee. There are no clarification requests or other signs of comprehension problems from the interlocutors.

References

Aelbrecht, Lobke, and Liliane Haegeman. 2012. VP-Ellipsis Is Not Licensed by VP-Topicalization. *Linguistic Inquiry* 43: 591–614. [CrossRef]

Allwood, Jens. 1976. The complex NP constraint as a non-universal rule and some semantic factors influencing the acceptability of Swedish sentences which violate the CNPC. In *University of Massachusetts Occasional Publications in Linguistics II*. Edited by Justine Stillings. Amherst: University of Massachusetts, pp. 1–20.

Allwood, Jens. 1999. The Swedish spoken language corpus at Göteborg University. In *Gothenburg Papers in Theoretical Linguistics*. Göteborg: Göteborg University, p. 81.

Andersson, Lars-Gunnar. 1975. *Form and Function of Subordinate Clauses*. (Gothenburg Monographs in Linguistics 1). Gothenburg: Department of Linguistics, University of Gothenburg.

Andersson, Lars-Gunnar. 1982. What is Swedish an exception to? In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabet Engdahl and Eva Ejerhed. Stockholm: Almqvist & Wiksell International, pp. 33–45.

Andersson, Sven-Gunnar, and Sigmund Kvam. 1984. *Satzverschränkung im Heutigen Deutsch. Volume 24 of Studien zur deutschen Grammatik*. Tübingen: Gunter Narr.

Bentzen, Kristine, and Merete Anderssen. 2019. The form and position of pronominal objects with non-nominal antecedents in Scandinavian and German. *Journal of Comparative Germanic Linguistics* 22: 169–188. [CrossRef]

Bentzen, Kristine, Jason Merchant, and Peter Svenonius. 2013. Deep properties of surface pronouns: Pronominal predicate anaphors in Norwegian and German. *Journal of Comparative Germanic Linguistics* 16: 97–125. [CrossRef]

Boersma, Paul, and David Weenink. 2020. Praat: Doing Phonetics by Computer [Computer Program, Version 6.1.16]. Available online: http://www.praat.org/ (accessed on 6 June 2020).

Bosch, Peter, Graham Katz, and Carla Umbach. 2007. The non-subject bias of German demonstrative pronouns. In *Anaphors in Text: Cognitive, Formal and Applied Approaches to Anaphoric Reference*. Edited by Monika Schwarz-Friese, Manfred Consten and Mareile Knees. Amsterdam: John Benjamins, pp. 154–64.

Bouma, Gerlof. 2008. Starting a Sentence in Dutch. Ph.D. thesis, University of Groningen, Groningen, The Netherlands.

Brandtler, Johan. 2019. The question of form in the forming of questions. The meaning and use of clefted *wh*-interrogatives in Swedish. *Journal of Linguistics* 55: 755–54. [CrossRef]

Chaves, Rui P., and Michael T. Putnam. 2020. *Unbounded Dependency Constructions: Theoretical and Experimental Perspectives*. Oxford: Oxford University Press.

Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*. Edited by Steven R. Andersson and Paul Kiparsky. New York: Holt, Rinehart and Winston, pp. 232–86.

Chung, Sandra, and James McCloskey. 1983. On the interpretation of certain island facts in GPSG. *Linguistic Inquiry* 14: 704–13.

Cinque, Guglielmo. 2013. *Typological Studies. Word Order and Relative Clauses*. New York: Routledge, pp. 218–22.
Daneš, František. 1974. Functional sentence perspective and the organization of the text. In *Papers on Functional Sentence Perspective*. Edited by Frantisek Daneš. Prague: Academia, pp. 106–28.
Engdahl, Elisabet. 1982. Restrictions on unbounded dependencies in Swedish. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabet Engdahl and Eva Ejerhed. Stockholm: Almqvist & Wiksell International, pp. 151–174.
Engdahl, Elisabet. 1997. Relative clause extractions in context. *Working Papers in Scandinavian Syntax* 60: 51–79.
Engdahl, Elisabet. 2012. Optional Expletive Subjects in Swedish. *Nordic Journal of Linguistics* 35: 99–144. [CrossRef]
Engdahl, Elisabet, and Annie Zaenen. 2020. Grammatical function selection in Swedish object shift. In *Proceedings of the LFG20 Conference, On-Line*. Edited by Miriam Butt and Ida Toivonen. Stanford: CSLI Publications, pp. 109–26.
Engdahl, Elisabet, and Eva Ejerhed, eds. 1982. *Readings on Unbounded Dependencies in Scandinavian Languages*. Stockholm: Almqvist & Wiksell International.
Engdahl, Elisabet, and Filippa Lindahl. 2014. Preposed object pronouns in mainland Scandinavian. *Working Papers in Scandinavian Syntax* 92: 1–32.
Erteschik-Shir, Nomi. 1973. On the Nature of Island Constraints. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Erteschik-Shir, Nomi. 2007. *Information Structure. The Syntax-Discourse Interface*. Oxford: Oxford University Press.
Faarlund, Jan Terje. 2019. *The Syntax of Mainland Scandinavian*. Oxford: Oxford University Press.
Fanselow, Gisebert. 2016. Syntactic and prosodic reflexes of information structure in Germanic. In *The Oxford Handbook of Information Structure*. Edited by Caroline Féry and Shinichiro Ishihara. Oxford: Oxford University Press.
Frey, Werner. 2004. A medial topic position for German. *Linguistische Berichte* 198: 153–90.
Frey, Werner. 2006. Contrast and movement to the German prefield. In *The Architecture of Focus*. Edited by Valéria Molnár and Susanne Winkler. Berlin/New York: Mouton de Gruyter, pp. 235–64.
Frey, Werner. 2007. Some contextual effects of aboutness topics in German. In *Interfaces and Interface Conditions*. Edited by Andreas Späth. Berlin/New York: Mouton de Gruyter, pp. 329–48.
Grice, Paul. 1975. Logic and Conversation. In *Speech Acts. Volume 3 of Syntax and Semantics*. Edited by Peter Cole and Jerry Morgan. New York: Academic Press.
Gundel, Jeanette K., and Thorstein Fretheim. 2004. Topic and focus. In *The Handbook of Pragmatics*. (Blackwell Handbooks in Linguistics 16). Edited by Laurence Horn and Gergory Ward. Malden: Blackwell, pp. 175–96.
Hankamer, Jorge, and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7: 391–428.
Hardt, Daniel. 1999. Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy* 22: 187–221. [CrossRef]
Holmberg, Anders. 2020. On the bottleneck hypothesis of verb second in Swedish. In *Rethinking Verb Second*. Edited by Rebecca Woods and Sam Wolfe. Oxford: Oxford University Press, pp. 40–60.
Jensen, Anne. 2002. Sætningsknuder i dansk [Knot sentences in Danish]. *NyS—Nydanske Studier & Almen Kommunikationsteori* 29: 105–24.
Johannessen, Janne Bondi, Joel Priestley, Kristin Hagen, Tor A. Åfarli, and Øystein A. Vangsnes. 2009. The Nordic Dialect Corpus—An advanced research tool. In *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009*. Edited by Kristiina Jokinen and Eckhard Bick. Odense: NEALT. pp. 73–80.
Johnson, Kyle. 2001. What VP Ellipsis Can Do, and What it Ca not, However, Not Why. In *The Handbook of Contemporary Syntactic Theory*. Edited by Mark Baltin and Chris Collins. Malden: Blackwell, pp. 439–79. [CrossRef]
Jörgensen, Nils. 1976. *Meningsbyggnaden i talad svenska [Sentence structure in spoken Swedish]*. Lund: Studentlitteratur.
Keenan, Edward L. 1987. A semantic definiton of "indefinite NP". In *The Representation of (In)definiteness*. Edited by Eric J. Reuland and Alice G. B. ter Meulen. Cambridge: MIT Press, pp. 286–317.
Kush, Dave, Akira Omaki, and Norbert Hornstein. 2013. Microvariation in islands? In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 239–64.
Kush, Dave, Charlotte Sant, and Sunniva Briså Strætkvern. 2021. Learning island-insensitivity from the input: A corpus analysis of child- and youth-directed text in Norwegian. *Glossa: A Journal of General Linguistics* 6: 105. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating variation in island effects. A case study of Norwegian *wh*-extraction. *Natural Language and Linguistic Theory* 36: 743–79. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95: 393–420. [CrossRef]
Lambrecht, Knud. 1988. There was a farmer had a dog: Syntactic amalgams revisited. In *Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Department of Linguistics, UCLA, vol. 14, pp. 319–39.
Lindahl, Filippa. 2010. Spetsställda led och Rematiska Relativer. En Korpusstudie av satsfläta med Presenteringsomskrivning/Existentialsats [A Corpus Study of Extraction from Presentational Clefts and Existential Constructions]. MA thesis, Department of Swedish, University of Gothenburg, Gothenburg, Sweden.
Lindahl, Filippa. 2011. Spetsställda bisatser i satsfläta med relativsats. [Preposed subordinate clauses in relative clause extraction]. *Språk och stil* 21: 199–204.
Lindahl, Filippa. 2017a. Det spetsställda ledets funktion i satsfläta med relativsats [The function of the preposed consituent in extraction from relative clauses]. *Norsk Lingvistisk Tidsskrift 35*: 49–66.

Lindahl, Filippa. 2017b. *Extraction from Relative Clauses in Swedish*. Gothenburg: Department of Swedish, University of Gothenburg. Available online: http://hdl.handle.net/2077/51985 (accessed on 1 March 2022).
Lindahl, Filippa. 2022. Extraction from relative clauses in Icelandic and Swedish: A parallel investigation. *Languages*. Submitted.
Lindahl, Filippa. to appear. Spetsställning och satsflätor i färöiskan. En översättningsstudie [Preposing and long-distance extraction in Faroese. A translation study]. In *Frændafundur 10. Proceedings of the Conference Frændafundur in Tórshavn 2019*. Edited by Bergur Rönne Moberg and Höskuldur Þráinsson. Tórshavn: Fróðskapur.
Lindahl, Filippa. under review. *The Prefield in Spoken Icelandic and Swedish*. Gothenburg: University of Gothenburg. Manuscript under review.
Lindahl, Filippa, and Elisabet Engdahl. Forthcoming. The pragmatics and syntax of pronoun preposing. A study of spoken Swedish. *Journal of Comparative Germanic Linguistics*.
Lutz, Ulli, Gereon Müller, and Arnim von Stechow, eds. 2000. *Wh-Scope Marking*. Amsterdam: John Benjamins.
Lødrup, Helge. 2012. Some Norwegian "Type Anaphora" are Surface Anaphora. *Journal of Germanic Linguistics* 24: 23–52. [CrossRef]
Löwenadler, John. 2015. Relative clause extraction: Pragmatic dominance, processing complexity and the nature of crosslinguistic variation. *Nordic Journal of Linguistics* 38: 37–65. [CrossRef]
Lühr, Rosemarie. 1988. Zur Satzverschränkung im heutigen Deutsch. *Groninger Arbeiten zur Germanistischen Linguistik* 29: 74–87.
McCawley, James D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53: 99–149. [CrossRef]
Mikkelsen, Kristian Mathias. 1894. *Dansk Sproglære med Sproghistoriske Tillæg. Haandbog for Lærere og Viderekomne [Danish Grammar]*. Copenhagen: Lehmann & Stage.
Mikkelsen, Line. 2015. VP anaphora and verb-second order in Danish. *Journal of Linguistics* 51: 595–643. [CrossRef]
Müller, Christiane, and Clara Ulrich Eggers. 2022. Island extractions in the wild: A corpus study of adjunct and relative clause islands in Danish and English. *Languages* 7: 125. [CrossRef]
Myrberg, Sara and Tomas Riad. 2015. The prosodic hierarchy of Swedish. *Nordic Journal of Linguistics* 38: 115–47. [CrossRef]
Ørsnes, Bjarne. 2011. Non-finite *do*-support in Danish. *Empirical Issues in Syntax and Semantics* 8: 409–34.
Ørsnes, Bjarne. 2013. VP anaphors and object shift: What do VP anaphors reveal about the licensing conditions for object shift in Danish? *Nordic Journal of Linguistics* 36: 245–74. [CrossRef]
Platzack, Christer. 2014. A feature driven account of mainland Scandinavian extraction from relative clauses. In *Talk Given at the Grammar Seminar at the Centre for Languages and Literature*. Lund: Lund University, April 3.
Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27: 53–94. [CrossRef]
Rizzi, Luigi. 1990. *Relativized Minimality*. (Linguistic Inquiry Monographs 16). Cambridge: The MIT Press.
Ross, John R. 1967. Constraints on Variables in Syntax. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Sigurðsson, Halldór Ármann, and Jim Wood. 2020. 'We Olaf": Pro[(x-)NP] constructions in Icelandic and beyond. *Glossa: A Journal of General Linguistics* 5: 16. [CrossRef]
Søfteland, Åshild. 2013. *Utbrytningskonstruksjonen i Norsk Spontantale [The Cleft Construction in Spoken Norwegian]*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
Sundman, Marketta. 1980. *Existentialkonstruktionen i svenskan [The existential construction in Swedish]*. (Meddelanden från stiftelsens för Åbo akademi forskningsinstitut 57). Åbo: The Research Institute of the Åbo Akademi Foundation.
Taraldsen, Knut Tarald. 1981. On the theoretical interpretation of a class of "marked" extractions. In *The Theory of Markedness in Generative Grammar*. Edited by Adriana Beletti, Luciana Brandi, and Luigi Rizzi. Pisa: Scuola Normale Superieore, pp. 475–516.
Teleman, Ulf, Staffan Hellberg, and Erik Andersson. 1999. *Svenska Akademiens Grammatik [Swedish Academy Grammar]*. Stockholm: Norstedts.
Theiler, Nadine and Gerlof Bouma. 2012. Two for the price of one: An LFG treatment of sentence initial object *es* in German. In *Proceedings of the LFG12 Conference*. Edited by Miriam Butt and Tracy Holloway King. Stanford: CSLI Publications, pp. 603–23.
Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.
Thráinsson, Höskuldur, Hjalmar P. Petersen, Jógvan í Lon Jacobsen, and Zakaris Svabo Hansen. 2004. *Faroese. An Overview and Reference Grammar*. Tórshavn: Føroya Fróðskaparfelag.
van Craenenbroeck, Jeroen and Liliane Haegeman. 2007. The derivation of subject initial V2. *Linguistic Inquiry* 38: 167–78. [CrossRef]
van Kampen, Jacqueline. 2007. Anaphoric pronouns for topic devices: Theoretical claims and acquisitional evidence. In Proceedings of the 2007 Child Language Seminar, Reading, UK, July 18–20; pp. 64–72.
Ward, Gregory. 1985. The Semantics and Pragmatics of Preposing (Topicalization). Ph.D. thesis, University of Philadelphia, Philadelphia, PA, USA.
Wellander, Erik. 1939. *Riktig svenska: En handledning i svenska språkets vård* [*Proper Swedish*]. Stockholm: Norstedt & Söner.
Zaenen, Annie. 1985. *Extraction Rules in Icelandic*. New York: Garland.
Zifonun, Gisela, Ludger Hoffman, and Bruno Strecker. 1997. *Grammatik der deutschen Sprache I–III*. Berlin: Walter de Gruyter.

languages

MDPI

Article

Extraction from Relative Clauses in Icelandic and Swedish: A Parallel Investigation

Filippa Lindahl [1,2]

[1] Division of Educational Science and Languages, University West, 46186 Trollhättan, Sweden; filippa.lindahl@hv.se
[2] Department of Swedish, University of Gothenburg, Box 200, 40530 Gothenburg, Sweden

Abstract: Extraction from relative clauses is generally taken to be unacceptable in Icelandic, unlike in the Mainland Scandinavian languages. Recent studies on Mainland Scandinavian show that the type of dependency as well as the embedding predicate matters for the acceptability of such extractions, and the study of spontaneously produced examples has improved our ability to create felicitous extraction contexts. The studies of Icelandic extraction predate these findings, and there is to date no study which systematically compares parallel sentences in Icelandic and Mainland Scandinavian. This article presents such a study, using two acceptability judgment experiments, one in Icelandic and one in Swedish, drawing on newly gained insights about fronting conditions in the two languages to create plausible contexts. The Icelandic participants rated extraction from relative clauses as unnatural, with a very large acceptability cost compared to in situ versions and good fillers. Extraction from *að*-clauses received mixed ratings, and local fronting was rated on a par with the in situ versions. In Swedish, extraction from relative clauses was rated as natural a majority of the time. There was no extraction cost in local fronting, extraction from *att*-clauses, or extraction from relative clauses in existential sentences, while extraction with other embedding predicates incurred some cost. No differences relating to the embedding predicate were seen in Icelandic. The study corroborates the view that extraction from relative clauses is unacceptable in Icelandic.

Keywords: A-bar movement; extraction; Icelandic; island phenomena; relative clauses; Scandinavian; Swedish; syntactic dependencies

Citation: Lindahl, Filippa. 2022. Extraction from Relative Clauses in Icelandic and Swedish: A Parallel Investigation. *Languages* 7: 163. https://doi.org/10.3390/languages7030163

Academic Editors: Anne Mette Nyvad and Ken Ramshøj Christensen

Received: 18 April 2022
Accepted: 8 June 2022
Published: 29 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Extraction from relative clauses is well attested in the Mainland Scandinavian languages, and studies have shown that such sentences involve grammatical A'-movement from a relative clause (e.g., Christensen and Nyvad 2014; Engdahl 1997; Lindahl 2014, 2017). A well known example from Swedish is given in (1).

(1) *De blommorna*$_1$ känner jag en man [som säljer _$_1$]. (Swe)
those flowers know I a man who sells
'I know a man who sells such flowers.' (Allwood 1976, p. 11)

These extractions were noted fairly early by Mainland Scandinavian grammarians working within descriptive or normative traditions (Mikkelsen 1894; Wellander 1939). In the '70s and '80s, Mainland Scandinavian extractions attracted attention in international syntax research due to work by Erteschik-Shir (1973); Allwood (1976); Engdahl and Ejerhed (1982), and others. Many theories of syntactic locality are specifically designed to exclude sentences like those in (1) (Chomsky 1964, 1973, 2001; Ross 1967). Important theoretical questions have thus been how to square the Mainland Scandinavian relative clause facts with theories of locality, and why certain languages permit this type of A'-dependencies while others, like English, German, and most other languages where it has been studied, do not.[1]

The Insular Scandinavian languages are of special interest for this typological question, since they share many but not all syntactic features with Mainland Scandinavian. Examples

parallel to (1) are unacceptable in both Icelandic and Faroese (Thráinsson et al. 2004; Zaenen 1985).

(2) a. * *Þessi blóm*$_1$ þekki ég mann [sem selur _$_1$] (Ice)
these flowers know I man who sells
(Zaenen 1985, p. 133)

b. * *Slíkar blómur*$_1$ kenni eg ein mann [sum selur _$_1$]. (Fa)
such flowers know I a man who sells
(Platzack 2014, p. 10)

This phenomenon is less well studied in Insular Scandinavian, however.[2] Zaenen's (1985) study, which provides an in-depth account of extraction rules in Icelandic, predates the recent wave of research on Mainland Scandinavian, where the knowledge of extraction from relative clauses has been advanced through large-scale acceptability experiments and the study of spontaneously produced examples in context, as well as through theoretical work and native speaker judgments (Christensen and Nyvad 2014; Engdahl 1997; Kush et al. 2018, 2019; Lindahl 2014, 2017; Nyvad et al. 2017). This research has taught us more about factors which affect the acceptability of extraction, such as the embedding predicate, the information-structural function of the fronted phrase, and the context the extraction sentence occurs in. Additionally, there is to date no acceptability study that compares extraction from relative clauses in Mainland Scandinavian and Insular Scandinavian using parallel example sentences. The type of extraction sentence that has been shown to be most common in Mainland Scandinavian—fronting of a topical pronoun from a relative clause in an existential sentence, see below—has to my knowledge not been discussed in the research on Insular Scandinavian.

The purpose of this article is therefore to study extraction from relative clauses in Icelandic, on the one hand, and in Swedish, on the other, building on insights from recent work. The study consist of an acceptability experiment in each language, where examples are kept as parallel as possible. The main aim is to obtain comparable data from the two languages, which will inform future typological and theoretical work. The article is structured as follows: Section 2 gives a brief overview of previous research with a focus on those factors that have been shown to ameliorate extraction in Mainland Scandinavian. Section 3 introduces the experiment, describing the methodology, the test sentences, and the participants. I then turn to the results in Section 4, where some clear differences between Swedish and Icelandic are shown. While the test sentences with extraction from relative clauses were mostly considered natural sounding by the Swedish participants, the Icelandic participants unanimously reject them, and factors that improve acceptability in Swedish do not seem to play any important role in Icelandic. Section 5 concludes with a discussion of the results and their implications.

2. Recent Research

This section briefly outlines recent research on extraction and fronting in Scandinavian relevant to the experimental design of the current study. Section 2.1 discusses the role of the embedding predicate, and Section 2.2 the fronted phrase and its relation to the context.

2.1. The Embedding Predicate

Early work established that the embedding predicate affects the acceptability of extraction from relative clauses. Erteschik-Shir (1973) observes that it is perceived as more acceptable to extract from a relative clause embedded under a predicate like *være* 'be/exist', *findes* 'exist', or *kende* 'know' in Danish, than from one embedded under a predicate like *pege på* 'point at', as illustrated in (3), where the judgments are Erstechik-Shir's.[3]

(3) a. *Det*$_1$ er der mange [der har gjort _$_1$]. (Da)
that is there many that have done
'There are many people who have done that.' (Erteschik-Shir 1973, p. 63)

b. *Det*$_1$ kender jeg mange [der har gjort _$_1$].
that know I many that have done
'I know many people who have done that.' (Erteschik-Shir 1973, p. 63)

c. * *Det*$_1$ har jeg peget paa mange [der har gjort _$_1$].
that have I pointed at many that have done
(Erteschik-Shir 1973, p. 64)

Erteschik-Shir attributes this to the pragmatic status of the relative clause in the utterance. If it is *dominant*, i.e., not presupposed or given, then extraction is more acceptable, on her account.[4] The information impact of the relative clause in context thus determines extraction possibilities.

Erteschik-Shir's observations with respect to the embedding predicates hold up in the other Mainland Scandinavian languages as well. However, later research has explored different ways to interpret them. Kush et al. (2013) suggest that extraction is in fact only possible with embedding verbs which select small clause complements, and that acceptable extraction actually involves a reanalysis of the relative clause as a small clause. Subsequent research, however, has argued against this view (Christensen and Nyvad 2014; Lindahl 2014, 2017; Müller 2015). Müller (2015) finds no significant difference between embedding predicates that select small clauses and those that do not, and Christensen and Nyvad (2014) find that there are differences in acceptability, but related to the frequency of the embedding predicate rather than to its ability to select a small clause complement. Data from spontaneous language use also show that the phenomenon is not restricted to small clause environments. Lindahl (2017) studied extraction in spoken Swedish, and while the study showed that extraction is clearly most common from presentational relatives with *vara* 'be/exist' as the embedding predicate, which could be accounted for on the small clause analysis, there are also examples that do not fit with this explanation. 13% of extraction sentences in my sample of spoken Swedish involved extraction from the complement of a lexical verb. Lexical verbs like *känna* 'know' and *se* 'see' would be expected on Kush et al.'s (2013) approach, seeing as they can select small clauses, whereas others are not amenable to this type of analysis. A few of the cases that would be unaccounted for are given in (4).

(4) a. *det*$_1$ vet jag många [som har fastnat i _$_1$] (Swe)
that know I many that have gotten stuck in
'I know of many people who have gotten stuck on that.' (Lindahl 2017, p. 90)

b. *Det*$_1$ har jag inte träffat någon [som gjort _$_1$]
that have I not met someone that done
'I've never met anyone who has done that.' (Lindahl 2017, p. 88)

c. *det*$_1$ stör jag mig på folk [som säger _$_1$]
that annoy I me on people that say
'People who say that annoy me.' (Lindahl 2017, p. 89)

The extractions above would not be expected on a small clause account, since neither of the embedding predicates selects a small clause.

Notably, pragmatic proposals like Erteschik-Shir's dominance condition cannot account for the observed range of data either. Even though most spontaneously produced examples do adhere to the condition, there are also examples where the main clause is clearly dominant, like (5) from Swedish.

(5) *Det*$_1$ beundrar jag folk [som klarar _$_1$ rent psykiskt], att bara
that admire I people that manage purely psychologically to just
vänta.
wait
'I admire people who can deal with that psychologically, to just wait.'

(Lindahl 2017, p. 89)

To show whether a clause is dominant, Erteschik-Shir (1973) uses a test called the "lie test". The test diagnoses whether it is possible to contradict a certain utterance or part of an utterance by saying *that is a lie*. The part of the sentence that can be an antecedent to *that* in *that is a lie* can be interpreted as dominant. A test modified for Swedish from Lindahl (2017) shows that the relative clause in examples like (5) cannot be interpreted as dominant.

(6) Speaker A:

Jag beundrar folk som klarar det rent psykiskt. (Swe)
I admire people that manage it purely psychologically
'I admire people who can deal with that psychologically.'

Speaker B:

a. Det stämmer inte, det gör du inte.
'That's not right, you don't.'

b. # Det stämmer inte, folk klarar inte det
'That's not right, people can't.'

(Lindahl 2017, p. 157)

As we can see, it is only possible to contradict the whole utterance, not the content of the embedded relative clause on its own. This means that the relative clause cannot be interpreted as dominant and that extraction should not be possible. Thus the extraction in (5) is a counterexample to Erteschik-Shir (1973)'s dominance condition. (For further discussion of pragmatic approaches, see Lindahl 2017, Chapter 5).

Corpus studies of written Norwegian and Danish have shown that extraction from relative clauses is clearly most common in existential environments in these languages as well (Kush et al. 2021; Müller and Eggers 2022). The state-of-the-art thus seems to be that extraction is more frequent when the relative clause occurs in certain environments (in existential/presentational sentences and as the complement of certain verbs). Formal acceptability studies also show that acceptability varies depending on the embedding verb (Christensen and Nyvad 2014), as observed more informally by Erteschik-Shir (1973) and others. However, the formal acceptability studies do not confirm the small clause hypothesis, and various counterexamples to both this hypothesis and those put forth relating to the pragmatics of the relative clause occur in spontaneous language. The point of this paper is not to provide a new analysis of this state of affairs. The fact that acceptability is related to the embedding predicate in Mainland Scandinavian is important, however, both in constructing the experimental materials, and in interpreting the data.

2.2. The Fronted Phrase

Much of the international research on extraction has focused on question formation, i.e., fronting of a *wh*-phrase, as in the example from Ross (1967) (7).

(7) * *Who* does Phineas know a girl [who is working with $_1$]? (Ross 1967, p. 124)

However, it was noted early in the research on the Mainland Scandinavian languages that such *wh*-extraction out of relative clauses is not at all common in these languages, and what typically occurs is fronting of topical DPs (Engdahl 1997; Erteschik-Shir 1973; Lindahl 2010, 2017). These observations fit well with formal acceptability studies by Kush et al. (2018, 2019), where topicalization was shown to lead to better acceptability ratings than *wh*-extraction in Norwegian.

A related point is that extraction from relative clauses is highly context-dependent. This is highlighted by pragmatic approaches such as that proposed by Erteschik-Shir (1973), and has been argued by Engdahl (1997) and Lindahl (2017), among others. The experimental work by Kush et al. (2019) confirms that acceptability ratings are significantly higher if the extraction sentence which is being judged is shown in a context.

The studies of spontaneously produced extraction sentences have furthermore not only revealed that fronting of topical DPs is most common. Studies of these naturally occurring examples have shown that it is particularly common to front pronominals. In Lindahl's 2017 study, 56% of all sentences with extraction from a relative clause in a spoken language data set involved a fronted pronominal. This is perhaps not surprising, given that fronted pronominal objects are common in Mainland Scandinavian quite generally (Engdahl and Lindahl 2014).

Further, in-depth studies of the function of pronoun fronting show that it has various discourse functions in Mainland Scandinavian, for example focus chaining, topic chaining, and contrast (Engdahl and Lindahl 2014; Erteschik-Shir 2007; Lindahl and Engdahl forthcoming). An overview of pronominal fronting related to extraction can be found in Engdahl and Lindahl (2022). For the purposes of this article, we should note that the most commonly used fronted phrase in all of the Mainland Scandinavian languages is the neuter pronoun *det* 'it/that' (Engdahl and Lindahl 2014, 2022). An example is given in (8), where the context sentence is rendered in English.

(8) alcoholism is not a disease however (Swe)

det$_1$ stör jag mej på folk [som säger _$_1$]
that annoy I me on people that say

'People who say that annoy me.' (Lindahl 2017, p. 89)

The antecedent of the pronoun is underlined. In this case, it is sentential. The pronoun is in what Erteschik-Shir (2007) calls a focus chain with the antedecent, since the content of the antencedent is all new, and introduced in the preceding utterance. See also the similar function of the local fronting in (9), which is from the *Nordic Dialect Corpus* (Johannessen et al. 2009).[5]

(9) a. *int:* tycker du det är roligt med små barn? (Swe)
think you it is fun with small children
'Do you think small children are fun?'

b. *s1:* ja *det*$_1$ tycker jag faktiskt _$_1$
yes it think I actually
'Yes, I actually think *so*.'

In studies of Mainland Scandinavian, Engdahl and Lindahl (2014) and Lindahl and Engdahl (forthcoming) found that focus chaining is the most common discourse function for a fronted pronoun, and that *det* with a sentential or VP antecedent is the most frequently fronted phrase.

Since there are no studies of spontaneously produced extraction from relative clauses in Icelandic, the corresponding data for this language are missing. However, Lindahl (forthcoming b) compared the use of the prefield in declaratives in spoken Icelandic and Swedish, also using the *Nordic Dialect Corpus*. While it turns out that objects are only very rarely fronted in spoken Icelandic, the study demonstrates that when non-subject arguments are fronted, the most common phrase is *það* 'it/that', which corresponds to *det* 'it/that' in Mainland Scandinavian. We see an example from the corpus in (10).

(10) a. *s1*: var þetta söngelsk ætt? (Ice)
was this song-loving family
'Was this a family that loves singing?

b. *s2*: nei *það*$_1$ held ég nú ekki _$_1$
no that think I now not
'No, I don't think so, really.'

While pronoun fronting serves many purposes in Mainland Scandinavian, only the most common type, focus chaining, occurred in the Icelandic part of the corpus. In fact, all of the Icelandic examples from the NDC involved pronouns in a focus chaining relationship.

2.3. Summary

To sum up, recent research shows that extraction from relative clauses is most common and most easily accepted in Mainland Scandinavian languages when the relative clause is existential/presentational, and with a few other embedding predicates. With respect to the fronted phrase, it has been shown that fronting of topical DPs is more common than fronting of *wh*-phrases, and that such extractions receive better acceptability ratings. Providing a context also ameliorates extraction. In Mainland Scandinavian, the previous research has shown that the same types of phrases that are commonly fronted in the local clause are extracted from relative clauses. We do not have access to spontaneously produced examples of extraction from Icelandic, but previous research on local fronting shows that object fronting in declaratives is used more rarely in this language. When it is used, however, the fronted phrase is usually the pronoun *það* 'it/that', and establishes a focus chain with the antecedent.

3. The Experiment

I ran two acceptability judgment studies that tested the acceptability of extraction from relative clauses in Icelandic and Swedish. The test sentences for the two languages were created in parallel in order to make the results as comparable as possible, and the two studies were carried out in the same way. The main goal was to find out how acceptable extraction from a relative clause is in Icelandic, compared to how acceptable it is in Swedish.

The test sentences, which are described in more detail in Section 3.1, used a few different embedding predicates, including existential sentences, to see if extraction from the relative clause in such clauses is more acceptable in Icelandic than the types that have been described in the previous literature on this language. Important clues can also come from comparing extraction from a relative clause to other types of extraction. Therefore a design with a number of different types of extraction was chosen: local fronting, extraction from an *að/att* ('that')-clause, extraction from a relative clause, and extraction from a *wh*-clause.

The studies were performed in the form of two questionnaires, one in Icelandic and one in Swedish, using the online survey tool Sunet Survey. For each test item, the context sentence was displayed in italics, and the test sentence in plain style, as shown in Figure 1.

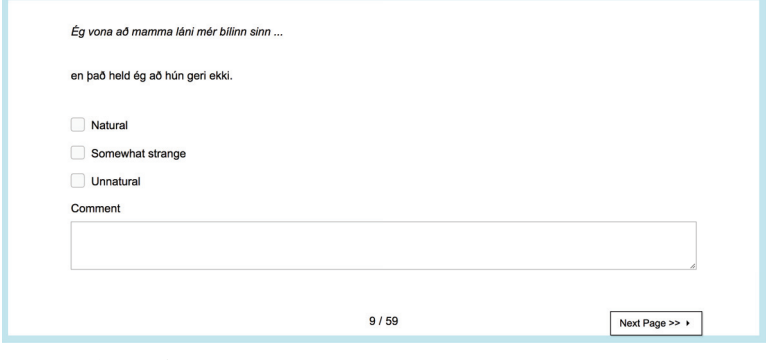


Figure 1. The interface in Sunet Survey.

The participants were asked to judge whether the test sentence was a natural follow-up in the context, using three answers: *natural*, *somewhat strange*, and *unnatural*. The questionnaire contained 52 test sentences, which are described in more detail in Section 3.1, and took 10–15 min to complete. The design builds on the assumption that participants will not rate ungrammatical or unacceptable test sentences as natural sounding. This simple experimental design, with only three possible answers and 52 test sentences, was chosen rather than e.g., the factorial design developed by Sprouse (2007), which has been used in many recent studies on extraction, due to the somewhat exploratory nature of the experiment. This makes direct comparisons between this study and others on Mainland

Scandinavian harder. However, since the main issue at hand here is whether there are differences between Icelandic and Swedish, the design is useful as it makes it relatively easy to set up two comparable experiments in the two languages. The participants and the distribution of the questionnaires are described in Section 3.2.

3.1. Test Sentences

The test sentences for the two languages were created at the same time, using as similar content as possible while maintaining the same syntactic structure, and making sure that the sentences sounded as natural as possible in both languages, apart from any strangeness or unacceptability associated with the fronting. The Swedish contexts and test sentences were judged to be natural sounding by the author and another native Swedish speaker. Two native Icelandic speakers helped construct the Icelandic sentences. An overview of the types of test sentences is given in (11)–(16), with examples from the Icelandic part of the experiment. Each sentence was presented in a context, shown as line one of the examples below. For each sentence with fronting, there was a corresponding test sentence with the pronoun in situ for comparison, in a similar context (the b-examples).[6]

(11) *Local fronting (2 sentence pairs)*

a. Margir segja að tungumálið hér hafi breyst ...
many say that language.DEF here has changed
'Many people say that the language here has changed ... '

en *það*$_1$ held ég ekki _$_1$.
but that think I not
'but I don't think so.'

b. Margir segja að veturnir hér séu orðnir kaldari ...
many say that winters.DEF here are become colder
'Many people say that the winter here have become colder ...'

en ég held *það* ekki.
but I think that not
'but I don't think so.'

(12) *Extraction from an* að*-clause (4 sentence pairs)*

a. Í gær tók Eiríkur hundinn sinn með sér á æfinguna ...
Yesterday took Eiríkur dog.DEF REFL.POSS with REFL at practice
'Yesterday, Eiríkur brought his dog to practice ...'

og *það*$_1$ held ég [að hann geri _$_1$ líka í dag].
and that think I that he does also today
'And I think he'll do that today too.'

b. Í gær tók Sveinn hundinn sinn með sér í vinnuna ...
Yesterday took Sveinn dog.DEF REFL.POSS with REFL to work.DEF
'Yesterday, Sveinn brought his dog to work ...'

og ég held [að hann geri *það* líka í dag].
and I think that he does that also today
'And I think he'll do that today too.'

(13) *Extraction from a relative clause (5 sentence pairs)*

a. Systir mín segir að við ættum að skipta yfir í sumartíma ...
sister mine says that we should to shift over on summer time
'My sister says that we should adopt daylight saving time ...'

en *það*$_1$ þekki ég engan annan [sem heldur _1].
but that know I no one else that thinks
'but I don't know anyone else who thinks so.'

b. Bróðir minn segir að við ættum að hafa evrur í stað króna ...
brother mine says that we should to have Euros in stead Icelandic króna
'My brother says that we should have Euros in stead of Icelandic króna ...'

en ég þekki engan annan [sem heldur *það*].
but I know no one else that thinks that
'but I don't know anyone else who thinks so.'

(14) *Extraction from a* wh*-clause (4 sentence pairs)*

a. Helgi og Gísli vilja fara á hátíð ...
Helgi og Gísli want go to festival
'Helge and Gísli want to go to a festival ...'

en *það*$_1$ er ég ekki viss [hvort þeir mega _1].
but that am I not sure whether they may
'but I'm not sure they are allowed.'

b. Bryndís og Erla vilja halda veislu ...
Bryndís and Erla want hold party
'Bryndís and Erla want to have a party ...'

en ég er ekki viss [hvort þær mega *það*].
but I am not sure whether they may that
'but I am not sure they are allowed.'

(15) *Good filler (10 sentences)*

Margir halda að túlípanar séu fallegri en rósir ...
many think that tulips are prettier than roses
'Many people think that tulips are prettier than roses ...'

en rósir eru hins vegar til í fleiri litum.
but roses are other ways to in more colors.
'but on the other hand, there are roses in more colors.'

(16) *Bad filler (12 sentences)*

Finnur sagði að við ættum að fara að synda fyrir kvöldmat ...
Finnur said that we should to go to swim before dinner
'Finnur said that we should go swimming before dinner ...'

og það ekki gerðist.
and that not happened

The Swedish test sentences are parallel. In total, there were 52 test sentences for each language (local/in situ: 2 sentence pairs, að-clause/in situ: 4 sentence pairs, rel. clause/in situ: 5 sentence pairs, *wh*-clause/in situ: 4 sentence pairs, good fillers: 10 sentences and bad fillers: 12 sentences).The complete list of test sentences can be found in Appendix A.

The test sentences start with a conjunction, either *en* 'but' or *og* 'and' (*men* and *och* in Swedish) to connect them to the context sentence. All of them used *það* 'it/that' (*det* 'it/that' in Swedish) as the fronted phrase, and the context was set up so that the pronominal had either a sentential antecedent or a VP antecedent, which is was in a focus chaining relation

to. This choice was based on the findings in previous research detailed in Section 2.2. The contexts were inspired by contexts from examples of spontaneously produced extraction sentences described by Engdahl (1997) and Lindahl (2017), as well as research on pronoun fronting (Engdahl and Lindahl 2014; Lindahl and Engdahl forthcoming; Lindahl forthcoming b).[7]

The five pairs of examples testing relative clause extraction used three different embedding predicates: *vera/vara* 'be/exist', specifically used in existential sentences in the experiment, and the transitive verbs *þekkja/känna* 'know', and *hitta/träffa* 'meet'. The verbs were chosen because they occur in spontaneously produced examples in Mainland Scandinavian. Since there are only a few test sentences with each verb, the choice to use different verbs was mainly to create variation in the test sentences. However, given what we know from previous research, it would not be surprising to see more positive judgments of the test sentences with *vara* 'be/exist' and somewhat worse, but still acceptable judgments with *känna* 'know' and *träffa* 'meet' in Swedish. Furthermore, if Icelandic were like Mainland Scandinavian, we would expect a similar pattern in this language. The heads of the relative clauses were quantified, using quantifiers like *margir/många* 'many', *enginn/ingen* 'no one'. This also builds on previous research: quantified heads are very common in spontaneously produced extraction sentences in Mainland Scandinavian (Engdahl 1997; Lindahl 2017).

The test sentences with *wh*-clauses include two pairs with embedded polar questions, and two pairs with embedded constituent questions. *Wh*-clauses permit extraction in both Mainland Scandinavian and Icelandic (Engdahl 1980; Zaenen 1985). Like the test sentences with extraction from relative clauses, the *wh*-clause test sentences were inspired by previous research. One difference between the Swedish and the Icelandic test sentences is that two of the pairs use the verb *undra* 'wonder' in Swedish, whereas the Icelandic version uses *ekki vera viss* 'not be sure', as this was the closest way to express the same thing while still using the same type of embedded polar question.

The test sentences with local fronting and *att/að*-clauses are included to provide data for comparison, to see if there is cost of fronting unrelated to crossing island inducing structures.[8]

The filler sentences, lastly, provide two baselines to compare the test sentences to. Both good fillers and bad fillers were also presented with a context sentence, and were designed to be similar to the test sentences in complexity. The good fillers are all grammatical, and could occur in everyday communication. The bad fillers have a grammatical context sentence, but all include some feature which makes them unacceptable. In (16), the unacceptability arises from the negation *ekki* 'not' preceding the finite verb, rendering an illicit V3 word order. The aim was to have a range of different types of errors, some very noticeable, like an unlicensed negative polarity expression, and some less stark, like a sentence involving embedded V2 in a context where this was not pragmatically licensed. It would be necessary to read the test sentences carefully to notice some of the errors. This means that the bad fillers also function as a control for whether the participants were paying attention or not.

The sentences were presented in a randomized order in the questionnaire. However, all participants saw the sentences in the same order. This means that there could potentially be some training effects, such that sentences seen later would receive better judgments (Christensen and Nyvad 2014). However, the exact same order of presentation was used in both Icelandic and Swedish, which should ensure that a comparison between the two languages is possible. See Table A5 for details on the order of presentation.

3.2. Participants

The questionnaire was distributed via Sunet Survey to students at the Department of Swedish at the University of Gothenburg and the Faculty of Icelandic and Comparative Cultural Studies at the University of Iceland. Some studies (Dąbrowska 2017; Schütze 1996) show that training in linguistics affects people's judgments in acceptability judgment tasks, which should be taken into account when interpreting the data. However, this selection of

participants has the advantage that the groups should be roughly comparable, which is crucial since the main point of the investigation is to compare extraction in Icelandic and Swedish.

The Swedish results are based on the answers of 27 native speakers who all reported that they spoke Swedish in their household growing up.[9] Two of the participants were bilingual from birth (Swedish-Spanish and Swedish-Turkish). All of the participants in the Swedish part of the study also reported speaking English, and several also speak one or more other languages, most often Spanish or French. The participants were between 21 and 58 years old (average: 29.5, median: 27). 19 were women, 7 men, and one non-binary.

29 participants answered the Icelandic questionnaire. All were native speakers, reporting that Icelandic was used in their household growing up. One of them was bilingual from birth (Icelandic-English). All of the Icelandic participants reported speaking English, and about half also some degree of Danish. Several also report speaking other languages, such as Spanish, French, or German. The participants were between 19 and 48 years old (average: 25, median: 24). 25 were women, 4 men.

4. Results

This section lays out the results of the study. The possible answers (*natural*, *somewhat strange*, and *unnatural*) are ordinal and the results for each test sentence can be expressed as a triplet, e.g., (5,10,5), where the first number gives the number of participants who chose the alternative *natural*, the second number *somewhat strange*, and the third number *unnatural*. For example, the triplet is (26,2,1) for the Icelandic good filler (17).

(17) ... en rósir eru hins vegar til í fleiri litum. (26,2,1)
but roses are other ways to in more colors.
'but on the other hand, there are roses in more colors.'

When an individual test sentence is discussed below, this triplet will be the measure of the acceptability reported for the sentence.

We can also calculate, for each sentence type, the percentage of the time each answer occurs across the test sentences of this type, e.g., how many times in total sentences with extraction from an *att*/*að*-clause were judged as *natural* etc. This will give a percentage for the sentence type for each answer.

4.1. Swedish

An overview of the calculation of answers per sentence type from the Swedish part of the study is given in Table 1.[10]

Table 1. Answers per sentence type, Swedish.

Sentence Type	Natural	Somewhat Strange	Unnatural
Local fronting	68.5%	29.6%	1.9%
In situ	55.6%	40.7%	3.7%
Extraction (*att*-clause)	77.8%	21.3%	0.9%
In situ	76.9%	21.3%	1.9%
Extraction (rel. clause)	57.0%	34.1%	8.9%
In situ	81.5%	15.6%	3.0%
Extraction (*wh*-clause)	65.7%	31.5%	2.8%
In situ	75.0%	22.2%	2.8%
Good fillers	77.0%	21.1%	1.9%
Bad fillers	8.6%	15.4%	75.9%

Interpreting the data, it should be noted that the participants have been rather strict in their judgments. The good fillers are all grammatical and were judged by the author and another native Swedish speaker as natural in context in preparation of the materials, but have only been judged as *natural* 77.0% of the time by the participants. 21.1% of the time, they were judged as *somewhat strange*, and 1.9% as *unnatural*. There are at least two potential reasons that the good fillers would not get 100% *natural* replies. First, answering what is natural is perhaps not an easy task for the participants, and they may not understand it the same way as the author. Second, the fact that the participants are students in the department of Swedish may matter. It is plausible that they—on a group level—are more attuned to style, clarity, and alternative ways to word a sentence than the author, who was focusing mainly on syntax and information structure.

Bad fillers were judged as *natural* 8.6% of the time, as *somewhat strange* 15.4% of the time, and as *unnatural* 75% of the time. It may seem surprising that bad fillers were considered *natural* to such a large extent. 8.6% equals 28 answers in absolute numbers. Upon further analysis, it turns out that 17 of these ratings were from two test sentences with a main clause question word order embedded under *att* 'that', as in (18).

(18) Min rumskompis vill måla vårt kök ljusgult ...
my roommate wants paint our kitchen bright yellow
'My roommate wants to paint our kitchen bright yellow ...'

men jag tror inte att kommer det att bli fint. (7,0,13)
but I think not that will it to become nice

This word order is usually not considered acceptable, but since it is fairly common to use main clause word order in some embedded contexts in Swedish (Teleman et al. 1999), a tendency that seems to be expanding, it may be the case that these are indeed acceptable to some of the participants. However, it should also be noted that flipping the order of *kommer* 'will' and *det* 'it', as in (19) renders the examples completely acceptable.

(19) men jag tror inte att det kommer att bli fint.
but I think not that it will to become nice
'but I don't think it will look good.'

It is possible that some participants read the example quickly and did not spot the problem. Another bad filler that received several *natural* judgments is given in (20).

(20) Banken förutspår att räntan stiger med 3 procent ...
bank.DEF foresees that interest.DEF rises with 3 percent
'The bank foresees that the interest rates will go up 3 percent ...'

och det går inte att hitta ekonomerna som vill uttala sig tydligare
and it goes not to find economists.DEF who want pronounce REFL clearer
än så. (5,10,12)
than so

To be more natural sounding, *ekonomerna* 'the economists' in this example should have been indefinite. However, this is a fairly minor change from the test sentence, and a fairly minor deviance in the first place, which might have led some participants to think the sentence is not that bad altogether.

4.1.1. Local Fronting

Turning to local fronting this was judged as *natural* 68.5% of the time, as *somewhat strange* 29.6% of the time, and as *unnatural* 1.9% of the time. This is better than the in situ versions, where the corresponding percentages were *natural*, 55.6%, *somewhat strange*, 40.7%, and *unnatural*, 3.7%. As discussed in detail in Lindahl and Engdahl (forthcoming), both of the word orders are grammatical in Swedish, and both occur in spontaneously produced

language. The fact that the version with fronting receives better judgments is not entirely unexpected: it is the more common version when the fronted pronoun is *det* with a VP or sentential antecedent, as in this case, and it fits well with the pragmatic strategies normally used in Swedish to start a sentence, since it involves focus chaining (Erteschik-Shir 2007; Lindahl and Engdahl forthcoming).

4.1.2. Extraction from *att*-Clauses

The Swedish participants' judgments for test sentences with extraction from an *att*-clause were very similar to their judgments for the good fillers. They picked the answer *natural* 77% of the time, *somewhat strange* 21.3% of the time, and *unnatural* only 0.9% of the time. The judgments for the in situ versions are almost identical: *natural* 76.9%, *somewhat strange* 21.3%, *unnatural* 1.9%. This means that we see no extraction cost for long extraction in a non-island environment. This could be because our method, with only three alternatives, allows less fine-grained distinctions. Another potential explanation would be that it is due to the type of fronted phrase we used. As we saw in the local fronting condition, the version with fronting actually received more *natural* judgments than the in situ version. It is conceivable that this effect counteracts an effect of similar size for long extraction, such that the acceptability cost of long extraction is hidden.

4.1.3. Extraction from Relative Clauses

Overall, the Swedish participants rated extraction from relative clause as *natural* over half of the time (57.0%). The rest of the time they mostly picked the alternative *somewhat strange* (34.1%). The answer *unnatural* was chosen only 8.9% of the time. These results are worse than the results for good fillers and extraction from an *att*-clause, but much better than for bad fillers. There is clearly a cost of extraction compared to the in situ versions, where the participants picked the answer *natural* 81.5% of the time, *somewhat strange* 15.6% of the time, and *unnatural* 3.0% of the time, which is fairly similar to the ratings for good fillers.

Looking closer at the individual test sentences, we find quite a bit of variation. Table 2 shows the relevant sentences. The context sentence is given in English, and the antecedent of the pronoun is underlined.

On one extreme, S15 on row 3 in the table, with extraction, is judged as natural 26 times out of 27. The judgments overall for this sentence are actually better than for the in situ version, S16, on row 4.[11] On the other extreme, the extraction sentence, S21, on row 9 received the judgment *natural* only ten times, and *unnatural* 6 times, which is much worse than the in situ version, S22, which was rated *natural* 24 times, and unnatural 0 times. The pair of sentences S13 and S14 on rows 1 and 2 are notable because the in situ version received worse ratings than the other in situ versions in that the option *somewhat strange* was chosen 12 times. Since both the context sentence and the test sentence are common and grammatical sentences, this most likely has something to do with the relation between the context sentence and the test sentence.[12] The version with extraction (11,14,2) was judged very similarly to the in situ version (12,12,3).

It comes as no surprise that the sentence S15 with *vara* 'be/exist' as the embedding predicate receives good ratings. It is also interesting to note that for both of the sentences with *vara*, S13 and S15, we essentially see no cost of extraction compared to the in situ version. For the sentences S17, S19, and S21 with the transitive verbs *känna* 'know' and *träffa* 'meet' as the embedding predicates, we see some extraction cost.

The relative clause extraction sentences were overall judged as somewhat less acceptable than the sentences with extraction from embedded questions, which will be discussed in the next section.

Table 2. Extraction from relative clauses: test sentences and judgments, Swedish.

Test Sentence		Natural	Somewhat Strange	Unnatural
S13 Olle claimed that it would rain men *det* var det ingen som trodde. *but that was there no one who thought*	extraction	11	14	2
S14 Anna claimed that it would snow men det var ingen som trodde *det*. *but there was no one who though that*	in situ	12	12	3
S15 You can use the gift card to buy a movie ticket och *det* är det många som gör. *and that are there many who do*	extraction	26	1	0
S16 You can use the wellness allowance to buy a gym card och det är många som gör *det*. *and there are many who do that*	in situ	23	4	0
S17 My sister says that we should switch to constant summer time men *det* känner jag ingen annan som tycker. *but that know I no one else who thinks*	extraction	15	9	3
S18 My brother says that we should have Euros instead of Kronor men jag känner ingen annan som tycker *det*. *but I don't know anyone else who thinks that*	in situ	25	2	0
S19 You can bike all the way to Riksgränsen men *det* har jag inte träffat någon som har gjort. *but that have I not met anyone who has done*	extraction	15	11	1
S20 You can hike all the way to Abisko men jag har inte träffat någon som har gjort *det*. *but I have not met anyone who has done that*	in situ	26	0	1
S21 My daughter wants to have a smart phone och *det* har jag träffat många andra barn som också vill. *and that have I met many other children who also want*	extraction	10	11	6
S22 My son wants to have his own computer och jag har träffat många andra barn som också vill *det*. *and I have met many other children who also want that*	in situ	24	3	0

4.1.4. Extraction from *wh*-Clauses

The Swedish test sentences for extraction from *wh*-clauses were rated as *natural* 65.7% of the time, as *somewhat strange* 31.5% of the time, and as *unnatural* 2.8% of the time. The in situ-versions received better ratings overall, at 75.0% *natural*, 22.2% *somewhat strange*, and 1.9% *unnatural*. There thus seems to be some cost of extraction from this type of clause as well, although smaller than the overall cost of extraction from a relative clause. We can also note that the ratings for the in situ versions are very close to the ones for the good fillers.

4.2. Icelandic

An overview of answers per sentence type in Icelandic is given in Table 3.

Table 3. Answers per sentence type, Icelandic.

Sentence Type	Natural	Somewhat Strange	Unnatural
Local fronting	67.2%	27.6%	5.2%
In situ	70.7%	24.1%	5.2%
Extraction (*að*-clause)	37.9%	29.3%	32.8%
In situ	97.4%	1.7%	0.9%
Extraction (rel. clause)	1.4%	6.9%	91.7%
In situ	84.8%	11.0%	4.1%
Extraction (*wh*-clause)	4.3%	32.8%	62.9%
in situ	69.8%	25.9%	4.3%
Good fillers	81.7%	15.2%	3.1%
Bad fillers	4.9%	14.4%	80.8%

Like the Swedish participants, the Icelandic participants have been rather strict in judging the good filler sentences. As mentioned, two native Icelandic speakers helped construct the Icelandic test sentences, which ought to be acceptable apart from any degradation related to the fronting. The participants judged the good fillers as *natural* 81.7% of the times, as *somewhat strange* 15.2% of the time, and as *unnatural* 3.1% of the time. The reasoning about the Swedish good filler sentences in Section 4.1 applies here too, and the fact that the results are quite close to the Swedish results for good fillers may indicate that the groups and test sentences are indeed fairly comparable, as hoped. With respect to the bad fillers the Icelandic participants chose the answer *natural* 4.9% of the time, *somewhat strange* 14.4% of the time, and *unnatural* 80.8% of the time, which is also quite close to the Swedish participants' judgments. A fairly large proportion of the *natural* and *somewhat strange* answers about the Icelandic bad fillers come from the same items, namely items I48 and I49, which are given in (21).

(21) a. Ég var á Akureyri í gær ...
I was in Akureyri yesterday
'I was in Akureyri yesterday ...'

og þar er hver einasta verslun sem gerir við hjól. (8,11,10)
and there is every store that fixes bikes

b. Ég var í Reykjavík í gær ...
I was in Reykjavík yesterday
'I was in Reykjavík yesterday ...'

og það er hver einasta verslun sem gerir við hjól þar. (1,15,13)
and EXPL is every store that fixes bikes there

The corresponding test sentences in Swedish received much worse judgments (0,4,23 and 0,2,25). The difference is likely related to the fact that Icelandic has an existential construction with universally quantified associates, unlike Swedish, i.e., that sentences like (22) are grammatical in Icelandic, but not in Swedish. See also (Milsark 1974; Thráinsson 2007).

(22) a. Það hafa allir kettirnir alltaf verið í eldhusinu. (Ice)
EXPL have all cats.DEF always been in kitchen.DEF
≈ 'All the cats have always been in the kitchen.' (Thráinsson 2007, p. 319)

b. * Det har alla katterna alltid varit i köket. (Swe)
EXPL have all cats.DEF always been in kitchen.DEF

4.2.1. Local Fronting

The Icelandic participants judged local fronting as *natural* 67.2% of the time, as *somewhat strange* 27.6% of the time, and as *unnatural* 5.2% of the time. The results for the in situ versions are very similar here. The participants chose the answer *natural* 70.7% of the time, *somewhat strange* 24.1% of the time, and *unnatural* 5.2% of the time. Both the fronting and in situ versions are rated worse than the good fillers, but there does not seem to be anything about the fronting in itself which makes the examples degraded, since the in situ versions received similar judgments.

4.2.2. Extraction from *að*-Clauses

The results for the sentences that tested extraction from *að*-clauses show a different pattern. Here, the versions with extraction were rated as *natural* only 37.9% of the time. 29.3% of the time they were rated as *somewhat strange*, and 32.7% of the time as *unnatural*. The in situ versions, on the other hand, were deemed natural sounding to a large extent. The participants judged the in situ version as *natural* 97.4% of the time, as *somewhat strange* 1.7% of the time, and as *unnatural* 0.9% of the time. There thus seems to be a large cost of extraction from an *að*-clause in Icelandic for this type of pronoun fronting.

4.2.3. Extraction from Relative Clauses

In extraction from relative clauses, we again see a very large cost of extraction. The participants rated the test sentences with extraction from a relative clause as *natural* only 1.4% of the time. The answer *somewhat strange* was chosen 6.9% of the time, and *unnatural* 91.7% of the time. In other words, the participants found extraction from relative clauses to be unnatural sounding across the board. The ratings are lower than for bad fillers. Furthermore, there does not seem to be anything wrong with the contexts or test sentences per se. The in situ versions were judged as *natural* 84.8% of the time, as *somewhat strange* 11.0% of the time, and as *unnatural* 4.1% of the time. These ratings are slightly higher than for the good fillers.

Turning to the individual test sentences, there is not much difference between them. The relevant sentences are shown in Table 4.

Table 4. Extraction from relative clauses: test sentences and judgments, Icelandic.

Test Sentence		Natural	Somewhat Strange	Unnatural
I13 Sindri claimed that it would rain en *því* var enginn sem trúði. *but that was no one who thought*	extraction	0	2	27
I14 Sigrún claimed that it would snow en það var enginn sem trúði *því*. *but there was no one who though that*	in situ	27	2	0
I15 You can use the gift card to buy a movie ticket og *það* eru margir sem gera. *and that are many who do*	extraction	2	7	20
I16 You can use the wellness allowance to buy a gym card og það eru margir sem gera *það*. *and there are many who do that*	in situ	27	2	0
I17 My sister says that we should switch to constant summer time en *það* þekki ég engan annan sem heldur. *but that know I no one else who thinks*	extraction	0	0	29
I18 My brother says that we should have Euros instead of Króna en ég þekki engan annan sem heldur *það*. *but I don't know anyone else who thinks that*	in situ	17	7	5
I19 You can bike all the way to Akureyrar en *það* hef ég ekki hitt neinn sem hefur gert. *but that have I not met anyone who has done*	extraction	0	0	29
I20 You can hike all the way to Keflavíkur en ég hef ekki hitt neinn sem hefur gert *það*. *but I have not met anyone who has done that*	in situ	28	1	0
I21 My daughter wants to have a smart phone og *það* hef ég hitt mörg önnur börn sem vilja líka. *and that have I met many other children who also want*	extraction	0	1	28
I22 My son wants to have his own computer og ég hef hitt mörg önnur börn sem vilja *það* líka. *and I have met many other children who also want that*	in situ	24	4	1

As the table reveals, ratings are grouped at the *unnatural* end of the scale for sentences with extraction, and in the *natural* end for in situ sentences, with few *somewhat strange* judgments across the board. Two test sentences deserve further comment. First, sentence I15, on row 3, was judged by two participants as *natural*, and seven participants judged it as *somewhat strange*. A clear majority rated it as *unnatural*, but it may be ever so slightly less unacceptable than the other sentences with extraction from a relative clause. The sentence I15 uses *vera* as the embedding predicate. However, the other sentence with *vera* and extraction, I13, is rated very poorly.

Second, sentence I18, stands out in receiving slightly worse judgments than the other in situ sentences. The participants have picked the option *somewhat strange* seven times, and *unnatural* five times. Most of the participants, 17, still rated the sentence as *natural*.

Overall, the sentences with extraction from relative clauses were not perceived as natural sounding by the Icelandic participants. As we will see in the next section, the results are similar for the sentences with extraction from *wh*-clauses.

4.2.4. Extraction from *wh*-Clauses

The Icelandic participants picked the alternative *natural* only 4.3% of the time judging the sentences with extraction from *wh*-clauses. The alternative *somewhat strange* was picked 32.8% of the time, and *unnatural* 62.9% of the time. A difference between these results and those for extraction from relative clauses is that there were more *somewhat strange*-answers, so extraction from *wh*-clauses appears not to be quite as unnatural as extraction from relative clauses to the Icelandic participants. Given the previous research (Zaenen 1985), the fact that these extractions received such low ratings is somewhat surprising. The results are also clearly worse than those for the sentences with extraction from *að*-clauses. It is worth pointing out that the in situ versions of the sentences also get somewhat worse results than good fillers and also than the in situ versions in extraction from relative clauses, which means that the test sentences may not have been entirely natural sounding to begin with. However, this can only account for some of the unnaturalness. I will discuss this further in Section 5.

4.3. Comparison of Icelandic and Swedish

Local fronting is rated quite similarly in Swedish and Icelandic. One difference, however, is that in Swedish, the fronted versions are overall better than the in situ versions, whereas in Icelandic, local fronting and in situ versions are on a par. In extraction from *att/að*-clauses, we see a clear difference between the two languages. In Swedish, the extraction sentences are on a par with the in situ-versions and with good fillers, but in Icelandic, these extractions seem to be somewhat degraded, while the in situ versions receive very favorable ratings. This can be illustrated with the examples in (23).

(23) a. men *det*$_1$ tror jag inte [att hon gör _$_1$]. (16,10,1) (Swe)
but that think I not that she does
'but I don't think she will.'

b. men jag tror inte [att han gör *det*]. (18,8,1)
but I think not that he does that
'but I don't think he will.'

c. en *það*$_1$ held ég [að hún geri ekki _$_1$]. (8,14,7) (Ice)
but that think I that she does not
'but I don't think that she will.'

d. en ég held [að hann geri *það* ekki]. (28,1,0)
but I think that he does that not
'but I don't think that he will.'

Overall, ratings are more degraded for extraction from *að*-clauses in Icelandic than from *att*-clauses in Swedish (Icelandic: 37.9%, 29.3%, 32.8% vs. Swedish: 77.8%, 21.3%, 0.9%).

Having looked more closely at ratings for comparable extraction sentences in Swedish and Icelandic, and using favorable pragmatic conditions, we are now in a position to compare the acceptability of extraction from relative clauses in the two languages, which was the overarching goal of this article. What we see is that controlling for factors relating to the embedding verb and the discourse function of the fronted phrase, there are clear differences between Swedish and Icelandic. In Swedish, extraction from relative clauses comes with some cost; extraction examples were overall rated worse than in situ versions, good fillers, and extraction from *att*-clauses. However, the majority of answers for the extraction sentences was still that it was natural sounding (57.0%, 34.1%, 8.9%). In Icelandic, ratings for the extraction sentences were very poor (1.4%, 6.9%, 91.7%), and the cost

compared to the in situ versions, extraction from *að*-clauses, and good fillers was much larger. Extraction from relative clauses was even rated worse than bad fillers. The test sentences in (24) illustrate the answer patterns in the two languages.

(24) Sindri claimed that it would rain ...

a. men *det*$_1$ var det ingen [som trodde _$_1$]. (11,14,2) (Swe)
but that was there no one who thought
'but there was no one who thought so.'

b. men det var ingen som trodde *det*. (12,12,3)
but there was no one who thought that
'but there was no one who thought so.'

c. en *því*$_1$ var enginn [sem trúði _$_1$]. (0,2,27) (Ice)
but that was no one who though
'but there was no one who though so.'

d. en það var enginn sem trúði *því*. (27,2,0)
but there was no one who thought that
'but there was no one who thought so.'

An interesting difference between the two languages is that in the Icelandic results, there is almost no variation in the judgments of the extraction sentences depending on the embedding predicate, unlike in Swedish. In Swedish, the extraction sentence and the in situ version got similar ratings in the examples where the embedding predicate was *vara* 'be/exist', i.e., in the existential sentences, but with the other embedding predicates, there was some extraction cost. In Icelandic, the extraction sentences are all rated poorly, and there is a large difference between the extraction version and the in situ version across all of the test sentences, the pattern looking the same regardless of the embedding predicate (See Table 4).

With respect to extraction from *wh*-clauses, these got much worse ratings in Icelandic than in Swedish (4.3%, 32.8%, 62.9% in Icelandic vs. 65.7%, 31.5%, 2.8% in Swedish). An example which illustrates the different answer patterns in the two languages is given in (25).

(25) There was only one person who could save the team from a loss ...

a. och *det*$_1$ visste alla [vem$_2$ det$_1$ var _$_2$] — Lionel Messi. (21,5,1) (Swe)
and that knew everyone who it was Lionel Messi
'and everyone knew who it was — Lionel Messi.'

b. og *það*$_1$ vissu allir [hver$_2$ _$_1$ var _$_2$] — Lionel Messi. (4,13,12) (Ice)
and that knew everyone who was Lionel Messi
'and everyone knew who it was — Lionel Messi.'

In Swedish, the trace of an extracted subject next to an overt element in the complementizer domain is spelled out as a resumptive pronoun, as we can see in (25a) (Engdahl 1985; Zaenen et al. 1981). In Icelandic, extraction of a subject could be expected to be acceptable, since Icelandic does not exhibit comp-trace effects (Zaenen 1985). However, as we see here, the ratings for this particular sentence were nevertheless poor.

Given these results, and the results for extraction from *að*-clauses and local fronting, it seems there is a potential difference between Swedish and Icelandic in how long-distance fronting of pronouns is treated. The type of pronoun fronting that was used seems to be acceptable in Icelandic in local fronting, but the fronting is often judged to be degraded out of embedded clauses, even *að*-clauses, which are not islands for movement in Icelandic generally.

5. Discussion

The goal of this article has been to investigate extraction from relative clauses in Icelandic and Swedish in parallel, building on insights from recent work on extraction in Mainland Scandinavian. The aim was to see what differences remain between the two languages when information structure and context are controlled for, and to find out if the same conditions that have proven to be favorable for extraction in Mainland Scandinavian would also improve Icelandic extraction sentences. The experiment revealed large differences between the two languages, corroborating previous work by Zaenen (1985); Thráinsson (2007), and others. Even though pronoun fronting of the kind used in the test sentences is acceptable in Icelandic, at least in local fronting, extraction from relative clauses and *wh*-clauses was judged to be unacceptable. This is different from the results from the Swedish part of the experiment, where the corresponding sentences were deemed *natural* in a majority of cases.

The results raise a question about long-distance fronting of pronouns in Icelandic, and under what discourse condition this type of movement is possible. In the study, only local fronting was judged as natural to a large extent by the Icelandic participants. These results are somewhat surprising given the previous research. Both extraction from *að*-clauses and extraction from *wh*-clauses are generally taken to be acceptable in Icelandic (Thráinsson 2007; Zaenen 1985). In relation to this, it is relevant to bring up contrast. There is a discussion in the previous literature about whether object fronting in Icelandic is only possible when the fronted phrase is contrasting with something (Light 2012; Lindahl forthcoming b). Lindahl (forthcoming b) argues that this is quite common in local pronominal fronting, but not necessary. Since the context sentences in the experiment were not set up to invoke a contrast, this may nevertheless have affected the Icelandic ratings, if contrast is the most common function for the fronted phrase in this language. In any case, it is intriguing that this mainly seems to affect the judgments of long-distance fronting but not local fronting. More research is clearly needed here. Comparing long-distance pronoun fronting with long-distance fronting of contrastive DPs and *wh*-phrases would be a natural next step.

From the perspective of theories of extraction, the results underscore that for all their similarities, the Scandinavian languages seem to be truly different when it comes to extraction from relative clauses. The conditions that are important in Mainland Scandinavian do not seem to play any role in Icelandic. That is, setting up a context which facilitates the type of pronoun fronting most commonly used in these languages, and using predicates that are known to facilitate extraction, does not lead to acceptable extraction from relative clauses in Icelandic.

Furthermore unlike in Swedish, no acceptability pattern related to the embedding predicate is visible. Contrast possibly plays a different role in Icelandic and could explain some of the difference in ratings between the two languages, and a future study could address this using contexts that evoke a contrastive interpretation of the fronted phrase. However, within Icelandic we still see a large difference between extraction from *að*-clauses and extraction from relative clauses. It thus seems likely that there is, in addition, some structural issue with extraction from relative clauses in this language.

The fact that I did not find an acceptability pattern related to the embedding predicate in the Icelandic part of the study warrants further comment. Engdahl and Lindahl (2022) report on a small study of Faroese indicating that the *sum*-clause in an existential sentence may permit extraction in this language.[13] The Faroese study used a different methodology than the current study, but if these results hold up in larger acceptability studies, Icelandic would be an extreme among the Scandinavian languages in not allowing extraction in this environment. Further acceptability studies using a more fine-grained scale in both Icelandic and Faroese would likely be enlightening.

Funding: This research was funded by an International Postdoc Grant (Dnr. 2017-06139) from the Swedish Research Council.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data reported in this study are available upon request.

Acknowledgments: I have presented parts of this work at the Grammar Seminar at the Department of Swedish at the University of Gothenburg and thank participants for helpful discussion of several points. I also want to thank Maia Andréasson, Elisabet Engdahl, Jóhannes Gísli Jónsson, Nick Kalivoda, and Sigríður Sigurjónsdóttir for discussion, help with materials, and finding participants for the study.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. Swedish test sentences.

	Test Sentence		
S1	Många säger att språket här har ändrat sig ... men det tycker jag inte.	local clause	extraction
S2	Många säger att vintrarna här har blivit kallare ... men jag tycker inte det.	local clause	in situ
S3	Anna säger att hon alltid vinner mot sin syster i schack ... och det gör hon också.	local clause	extraction
S4	Olle säger att han alltid vinner mot sin bror i tennis ... och han gör det också.	local clause	in situ
S5	Jag hoppas att min mamma ska låna ut sin bil ... men det tror jag inte att hon gör.	*att*-clause	extraction
S6	Jag hoppas att min bror ska låna ut sin motorcykel ... men jag tror inte att han gör det.	*att*-clause	in situ
S7	Jag tycker att sommaren är den bästa årstiden ... och det tror jag att de flesta håller med om.	*att*-clause	extraction
S8	Jag tycker att vintern är finast i december ... och jag tror att de flesta håller med om det.	*att*-clause	in situ
S9	Linas kollegor blir irriterade när hon kommer för sent ... men det tror jag inte att hon förstår.	*att*-clause	exrtaction
S10	Aminas klasskamrater blir sura när hon tar det sista kaffet ... men jag tror inte att hon förstår det.	*att*-clause	in situ
S11	Igår tog Erik med sin hund till träningen ... och det tror jag att han gör idag också.	*att*-clause	extraction
S12	Igår tog Sven med sin hund till jobbet ... och jag tror att han gör det idag också.	*att*-clause	in situ
S13	Olle påstod att det skulle regna ... men det var det ingen som trodde.	rel. clause	extraction
S14	Anna påstod att det skulle snöa ... men det var ingen som trodde det.	rel. clause	in situ
S15	Man kan använda presentkortet till att köpa en biobiljett ... och det är det många som gör.	rel. clause	extraction
S16	Man kan använda friskvårdsbidraget till att köpa ett gymkort ... och det är många som gör det.	rel. clause	in situ
S17	Min syster säger att vi borde gå över till ständig sommartid ... men det känner jag ingen annan som tycker.	rel. clause	extraction
S18	Min bror säger att vi borde ha euro istället för kronor ... men jag känner ingen annan som tycker det.	rel. clause	in situ

Table A1. *Cont.*

	Test Sentence		
S19	Man kan cykla hela vägen till Riksgränsen ... men det har jag inte träffat någon som har gjort.	rel. clause	extraction
S20	Man kan vandra hela vägen till Abisko ... men jag har inte träffat någon som har gjort det.	rel. clause	in situ
S21	Min dotter vill ha en smartphone ... och det har jag träffat många andra barn som också vill.	rel. clause	extraction
S22	Min son vill ha en egen dator ... och jag har träffat många andra barn som också vill det.	rel. clause	in situ
S23	Johan och Mikael tänker åka på festival ... men det undrar jag om de får.	*wh*-clause	extraction
S24	Karin och Tove tänker ha fest ... men jag undrar om de får det.	*wh*-clause	in situ
S25	Båtägarna ska städa området noggrant varje vecka ... men det undrar jag om de gör.	*wh*-clause	extraction
S26	Delägarna borde stå för kostnaden ... men jag undrar om de gör det idag.	*wh*-clause	in situ
S27	Någon har nyligen räknat ut hur man kan förutsäga jordbävningar ... och det är jag väldigt nyfiken på vem som har gjort.	*wh*-clause	extraction
S28	Någon har visst räknat ut hur man kan förutsäga översvämningar ... och jag är väldigt nyfiken på vem som har gjort det	*wh*-clause	in situ
S29	Det fanns bara en som kunde rädda laget från förlust ... och det visste alla vem det var - Lionel Messi.	*wh*-clause	extraction
S30	Det fanns bara en som kunde rädda landslaget från förlust ... och alla visste vem det var - Zinedine Zidane.	*wh*-clause	in situ

Table A2. Swedish filler sentences.

	Test Sentence	
S31	Många tycker att tulpaner är vackrare än rosor ... men rosor finns å andra sidan i flera nyanser.	Good filler
S32	Många tycker att rött vin är godare än vitt vin ... men vitt vin passar ändå bäst till fisk.	Good filler
S33	Min moster brukar vilja låna böcker på biblioteket ... så därför går vi dit tillsammans varje söndag.	Good filler
S34	Min faster brukar vilja se allt som ställs ut på konsthallen ... så därför följer jag med henne dit rätt ofta.	Good filler
S35	Det var många som trodde att Brasilien skulle vinna senaste fotbolls-VM ... men de spelade inte tillräckligt bra.	Good filler
S36	Det var många som trodde att Kanada skulle spela bra i senaste hockey-VM ... men de överträffade alla förväntningar.	Good filler
S37	Det är sällan man träffar folk som har varit i Kiruna ... men det händer ändå då och då.	Good filler
S38	Det är inte ofta man träffar folk som har varit vid Riksgränsen ... men jag har stött på några stycken i alla fall.	Good filler
S39	Iris kusin bjöd med mig på vargsafari i förra veckan ... och det var väldigt spännande.	Good filler
S40	Evas kusin bjöd med mig på bio i fredags ... och det var verkligen trevligt.	Good filler
S41	Isak sa att vi borde gå och simma innan middagen ... och så blev inte det.	Bad filler
S42	Samira sa att vi måste göra allt för att vinna orienteringstävlingen ... men inte blev det.	Bad filler
S43	Vi har bjudit in nästan hundra personer till festen ... och jag undrar verkligen vilka kommer.	Bad filler

Table A2. *Cont.*

	Test Sentence	
S44	Vi har bjudit in kanske två hundra personer till utställningen ... så jag tror att så värst många kommer.	Bad filler
S45	Min kollega ska presentera ett nytt gränssnitt imorgon ... och jag tror verkligen att kommer det bli bra.	Bad filler
S46	Min rumskompis vill måla vårt kök ljusgult ... men jag tror inte att kommer det att bli fint.	Bad filler
S47	Jag var hos min bror igår ... han bor nära mig, så jag ofta träffar honom.	Bad filler
S48	Jag var i Göteborg igår ... och där finns det varje affär som lagar cyklar.	Bad filler
S49	Jag var i Stockholm igår ... och det finns varje affär som lagar cyklar där.	Bad filler
S50	Man säger att det är 95 procent säkert ... och tydligare än så går det inte att hitta professorn som vill uttala sig.	Bad filler
S51	Banken förutspår att räntan stiger med 3 procent ... och det går inte att hitta ekonomerna som vill uttala sig tydligare än så.	Bad filler
S52	Vattnet är varmare i den andra poolen ... och Anna vill där oftast bada.	Bad filler

Table A3. Icelandic test sentences.

	Test Sentence		
I1	Margir segja að tungumálið hér hafi breyst ... en það held ég ekki.	local clause	extraction
I2	Margir segja að veturnir hér séu orðnir kaldari ... en ég held það ekki.	local clause	in situ
I3	Anna segir að hún vinni systur sína alltaf í skák ... og það gerir hún líka.	local clause	extraction
I4	Einar segir að hann vinni bróður sinn alltaf í tennis ... og hann gerir það líka.	local clause	in situ
I5	Ég vona að mamma láni mér bílinn sinn ... en það held ég að hún geri ekki.	*að*-clause	extraction
I6	Ég vona að bróðir minn láni mér mótorhjólið sitt ... en ég held að hann geri það ekki.	*að*-clause	in situ
I7	Ég held að sumarið sé besti tími ársins ... og því held ég að flestir séu sammála.	*að*-clause	extraction
I8	Ég held að veturinn sé fallegastur í desember ... og ég held að flestir séu sammála því.	*að*-clause	in situ
I9	Samstarfsmenn Línu verða pirraðir þegar hún kemur of seint ... en því held ég að hún átti sig ekki á.	*að*-clause	extraction
I10	Bekkjarfélagar Önnu verða reiðir þegar hún klárar kaffið ... en ég held að hún átti sig ekki á því.	*að*-clause	in situ
I11	Í gær tók Eiríkur hundinn sinn með sér á æfinguna ... og það held ég að hann geri líka í dag.	*að*-clause	extraction
I12	Í gær tók Sveinn hundinn sinn með sér í vinnuna ... og ég held að hann geri það líka í dag.	*að*-clause	in situ
I13	Sindri hélt því fram að það myndi rigna ... en því var enginn sem trúði.	rel. clause	extraction
I14	Sigrún hélt því fram að það myndi snjóa ... en það var enginn sem trúði því.	rel. clause	in situ
I15	Þú getur notað gjafakortið til að kaupa bíómiða ... og það eru margir sem gera.	rel. clause	extraction
I16	Þú getur notað styrkinn til að kaupa líkamsræktarkort ... og það eru margir sem gera það.	rel. clause	in situ

Table A3. *Cont.*

	Test Sentence		
I17	Systir mín segir að við ættum að skipta yfir í sumartíma ... en það þekki ég engan annan sem heldur.	rel. clause	extraction
I18	Bróðir minn segir að við ættum að hafa evrur í stað króna ... en ég þekki engan annan sem heldur það.	rel. clause	in situ
I19	Það er hægt að hjóla alla leið til Akureyrar ... en það hef ég ekki hitt neinn sem hefur gert.	rel. clause	extraction
I20	Það er hægt að ganga alla leið til Keflavíkur ... en ég hef ekki hitt neinn sem hefur gert það.	rel. clause	in situ
I21	Dóttir mín vill fá snjallsíma ... og það hef ég hitt mörg önnur börn sem vilja líka.	rel. clause	extraction
I22	Sonur minn vill fá tölvu ... og ég hef hitt mörg önnur börn sem vilja það líka.	rel. clause	in situ
I23	Helgi og Gísli vilja fara á hátíð ... en það er ég ekki viss hvort þeir mega.	*wh*-clause	extraction
I24	Bryndís og Erla vilja halda veislu ... en ég er ekki viss hvort þær mega það.	*wh*-clause	in situ
I25	Bátaeigendunum ber að þrífa svæðið vandlega í hverri viku ... en það er ég ekki viss hvort þau gera.	*wh*-clause	extraction
I26	Meðeigendurnir ættu að bera kostnaðinn ... en ég er ekki viss hvort þau gera það.	*wh*-clause	in situ
I27	Einhver hefur nýlega búið til tæki til að spá fyrir um jarðskjálfta ... og það er ég mjög forvitin(n) um hver gerði.	*wh*-clause	extraction
I28	Einhver hefur nýlega búið til tæki til að spá fyrir um flóð ... og ég er mjög forvitin(n) um hver gerði það.	*wh*-clause	in situ
I29	Það var bara einn sem gat bjargað liðinu frá tapi ... og það vissu allir hver var - Lionel Messi.	*wh*-clause	extraction
I30	Það var bara einn sem gat bjargað landsliðinu frá tapi ... og allir vissu hver það var - Zinedine Zidane.	*wh*-clause	in situ

Table A4. Icelandic filler sentences.

	Test Sentence	
I31	Margir halda að túlípanar séu fallegri en rósir ... en rósir eru hins vegar til í fleiri litum.	Good filler
I32	Margir halda að rauðvín sé betra en hvítvín ... en hvítvín hentar samt betur með fiski.	Good filler
I33	Frænka mín vill venjulega fá lánaðar bækur á bókasafninu ... svo við förum þangað saman á hverjum föstudegi.	Good filler
I34	Frænka mín vill venjulega sjá allt sem er til sýnis í listasafninu ... svo ég fer með henni þangað nokkuð oft.	Good filler
I35	Það voru margir sem héldu að Brasilíumenn myndi vinna síðasta heimsmeistaramót ... en þeir spiluðu ekki nógu vel.	Good filler
I36	Það voru margir sem héldu að Kanadamenn myndi spila illa á síðasta heimsmeistaramóti ... en þeir fóru fram úr öllum væntingum.	Good filler
I37	Það er sjaldgæft að hitta fólk sem hefur farið til Svalbarða ... en það gerist samt af og til.	Good filler
I38	Það er ekki oft sem maður hittir fólk sem hefur farið til Álandseyja ... en ég hef alla vega rekist á nokkra.	Good filler
I39	Frænka Guðrúnar bauð mér í hvalaskoðun í síðustu viku ... og það var mjög spennandi.	Good filler
140	Frænka Evu bauð mér í bíó síðasta föstudag ... og það var mjög skemmtilegt.	Good filler
I41	Finnur sagði að við ættum að fara að synda fyrir kvöldmat ... og það ekki gerðist.	Bad filler

Table A4. *Cont.*

	Test Sentence	
I42	Helga sagði að við yrðum að gera allt til að vinna ratleikinn ... en ekki það gerðist.	Bad filler
I43	Við höfum boðið næstum hundrað manns í veisluna ... og ég velti því fyrir mér hverjir sem að koma.	Bad filler
I44	Við höfum boðið um tvö hundruð manns á sýninguna ... en ég held að neinir komu.	Bad filler
I45	Samstarfsmaður minn mun kynna nýtt viðmót á morgun ... og ég held virkilega að verði það gott.	Bad filler
I46	Sambýlismaður minn vill mála eldhúsið okkar ljósgult ... en ég held að verði það ekki fallegt.	Bad filler
I47	Ég var með bróður mínum í gær. Hann býr nálægt mér ... svo ég oft hitti hann.	Bad filler
I48	Ég var á Akureyri í gær ... og þar er hver einasta verslun sem gerir við hjól.	Bad filler
I49	Ég var í Reykjavík í gær ... og það er hver einasta verslun sem gerir við hjól þar.	Bad filler
I50	Það er sagt að það sé 95 prósent öruggt ... og skýrar um málið er ekki til prófessorinn sem tjái sig.	Bad filler
I51	Bankinn spáir því að vextirnir hækki um 3 prósent ... og það er ekki til hagfræðingurinn sem tjái sig skýrar um málið.	Bad filler
I52	Vatnið er heitara í hinni lauginni ... og Anna vill þar venjulega synda.	Bad filler

Table A5. Order of presentation.

Test Sentence
S35/I35
S05/I05
S42/I42
S39/I39
S13/I13
S32/I32
S46/I46
S12/I12
S37/I37
S16/I16
S28/I28
S50/I50
S10/I10
S21/I21
S03/I03
S51/I51
S08/I08
S40/I40
S49/I49
S24/I24
S11/I11
S38/I38
S47/I47
S17/I17
S34/I34
S20/I20
S45/I45
S04/I04
S26/I26
S48/I48
S25/I25
S09/I09
S31/I31
S29/I29
S06/I06
S52/I52
S30/I30
S19/I19
S01/I01

Table A5. *Cont.*

Test Sentence
S33/I33
S43/I43
S36/I36
S15/I15
S41/I41
S18/I18
S27/I27
S02/I02
S44/I44
S14/I14
S07/I07
S22/I22
S23/I23

Notes

1 Although see Vincent (2021) and Vincent et al. (2022) for recent work on English which complicates the picture somewhat.

2 But see Engdahl and Lindahl (2022) and Lindahl (forthcoming a) for recent work on Faroese, where acceptable extraction seems to be limited to existential contexts.

3 I am following Erteschik-Shir in glossing *det* as *that* in English. However, *det* is ambiguous in writing in both Danish and Swedish. When unstressed, these pronouns are interpreted as personal pronouns, while the stressed versions function as distal demonstratives (Faarlund 2019, p. 27). I have systematically used *that* in all the glosses for fronted pronominal objects in this article, since it is not possible to tell whether the pronoun is stressed in constructed examples. However, given discussion in Engdahl and Lindahl (2014) and Engdahl and Lindahl (2022), both stressed and unstressed fronted pronouns often occur in spontaneous language use, so the fronted pronouns could just as well be interpreted as personal pronouns. I have done the same with the glosses for Icelandic *það* 'that/it', see below.

4 For more recent proposals along the same lines, see Van Valin (1994) and Goldberg (2006).

5 *int* stands for *interviewer* and *s1* for *speaker 1*.

6 Indices, gaps, and italics on the fronted phrase have been added here, but were not present in the questionnaires.

7 See also Engdahl and Lindahl (2022).

8 In the experiment, personal pronouns were used for the subjects of the embedded clauses in most cases, as this is common in everyday speech. A reviewer points out that the test sentences might have been more comparable if the subjects of the *att/að*-clauses were quantified, like the heads of the relative clauses in the relative clause test sentences. This insight could be incorporated in future work.

9 The total number of participants who answered the questionnaire were 31, but 4 non-native speakers were excluded from the analysis.

10 Due to the reduced experimental setup, the reporting is limited to descriptive statistics, showing percentages for the sentence types and the raw numbers for individual sentences.

11 A reviewer points out that there may be a training effect here, since S15 was presented late in the experiment.

12 One possibility is that some participants would have preferred *tro på* 'believe in' instead of *tro* 'believe' in the test sentences together with the verb *påstå* 'claim' in the context sentence.

13 For a more detailed account, see Lindahl (forthcoming a). See also McCawley (1981); Vincent (2021), and Vincent et al. (2022) who show that extraction is facilitated in English in existential and predicate nominal environments.

References

Allwood, Jens. 1976. The complex NP constraint as a non-universal rule and some semantic factors influencing the acceptability of Swedish sentences which violate the CNPC. In *University of Massachusetts Occasional Publications in Linguistics II*. Edited by Justine Stillings. Amherst: University of Massachusetts, pp. 1–20.

Chomsky, Noam. 1964. *Current Issues in Linguistic Theory*. Janua Linguarum. Series Minor 38. The Hague: Mouton.

Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*. Edited by Stephen R. Andersson and Paul Kiparsky. New York: Holt, Rinehart and Winston, pp. 232–86.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Language*. Edited by Michael Kenstowicz. Cambridge: MIT Press, pp. 1–52.

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2014. On the nature of escapable relative islands. *Nordic Journal of Linguistics* 37: 29–45.

Dąbrowska, Ewa. 2017. Naive v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review* 17: 1–23.

Engdahl, Elisabet. 1980. The Syntax and Semantics of Questions in Swedish. Ph.D. thesis, University of Massachusetts Amherst, Amherst, MA, USA.

Engdahl, Elisabet. 1985. Parasitic gaps, resumptive pronouns, and subject extractions. *Linguistics* 23: 3–44.
Engdahl, Elisabet. 1997. Relative clause extractions in context. *Working Papers in Scandinavian Syntax* 60: 51–79.
Engdahl, Elisabet, and Eva Ejerhed, eds. 1982. *Readings on Unbounded Dependencies in Scandinavian Languages*. Stockholm: Almqvist & Wiksell International.
Engdahl, Elisabet, and Filippa Lindahl. 2014. Preposed object pronouns in mainland Scandinavian. *Working Papers in Scandinavian Syntax* 92: 1–32.
Engdahl, Elisabet, and Filippa Lindahl. 2022. Extraction and pronoun preposing in Scandinavian. *Languages* 7: 128.
Erteschik-Shir, Nomi. 1973. On the Nature of Island Constraints. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Erteschik-Shir, Nomi. 2007. Information Structure. In *The Syntax-Discourse Interface*. Oxford: Oxford University Press.
Faarlund, Jan Terje. 2019. *The Syntax of Mainland Scandinavian*. Oxford: Oxford University Press.
Goldberg, Adele E. 2006. Constructions at Work. In *The Nature of Generalization in Language*. New York: Oxford University Press.
Johannessen, Janne Bondi, Joel Priestley, Kristin Hagen, Tor A. Åfarli, and Øystein A. Vangsnes. 2009. The Nordic Dialect Corpus—An advanced research tool. Paper presented at the 17th Nordic Conference of Computational Linguistics NODALIDA 2009, Odense, Denmark, May 14–16. Edited by Kristiina Jokinen and Eckhard Bick. Odense: NEALT, pp. 73–80.
Kush, Dave, Akira Omaki, and Norbert Hornstein. 2013. Microvariation in islands? In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 239–64.
Kush, Dave, Charlotte Sant, and Sunniva Briså Strætkvern. 2021. Learning island-insensitivity from the input: A corpus analysis of child- and youth-directed text in norwegian. *Glossa: A Journal of General Linguistics* 6: 1–50.
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating variation in island effects. A case study of Norwegian wh-extraction. *Natural Language and Linguistic Theory* 36: 743–79.
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95: 393–420.
Light, Caitlin. 2012. The Syntax and Pragmatics of Fronting in Germanic. Ph. D. thesis, University of Pennsylvania, Philadelphia, PA, USA.
Lindahl, Filippa. 2010. Spetsställda led och rematiska relativer. En korpusstudie av satsfläta med presenteringsomskrivning/existentialsats [A Corpus Study of Extraction from Presentational Clefts and Existential Constructions]. Master's thesis, Department of Swedish, University of Gothenburg, Gothenburg, Sweden.
Lindahl, Filippa. 2014. Relative clauses are not always strong islands. *Working Papers in Scandinavian Syntax* 93: 1–25.
Lindahl, Filippa. 2017. *Extraction from Relative Clauses in Swedish*. Gothenburg: Department of Swedish, University of Gothenburg.
Lindahl, Filippa, and Elisabet Engdahl. Forthcoming. The pragmatics and syntax of pronoun preposing. A study of spoken Swedish. *Journal of Comparative Germanic Linguistics*. *to appear*.
Lindahl, Filippa. Forthcoming a. Spetsställning och satsflätor i färöiskan. En översättningsstudie [Preposing and long-distance extraction in Faroese. A translation study]. Frændafundur 10. Paper presented at the Conference Frændafundur in Tórshavn 2019, Thorshofn, Iceland, August 16–18. Edited by Bergur Rönne Moberg and Höskuldur Þráinsson. Tórshavn: Fróðskapur. *to appear*.
Lindahl, Filippa. Forthcoming b. *The Prefield in Spoken Icelandic and Swedish.* Gothenburg: University of Gothenburg. *under review*.
McCawley, James D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53: 99–149.
Mikkelsen, Kristian Mathias. 1894. Dansk Sproglære med sproghistoriske Tillæg. In *Haandbog for Lærere og Viderekomne [Danish Grammar]*. Copenhagen: Lehmann & Stage.
Milsark, Gary. 1974. Existential Sentences in English. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Müller, Christiane. 2015. Against the small clause hypothesis: Evidence from Swedish relative clause extractions. *Nordic Journal of Linguistics* 38: 67–92.
Müller, Christiane, and Clara Ulrich Eggers. 2022. Island extractions in the wild: A corpus study of adjunct and relative clause islands in Danish and English. *Languages* 7: 125.
Nyvad, Anne Mette, Ken Ramshøj Christensen, and Sten Vikner. 2017. CP-recursion in Danish: A cP/CP analysis. *The Linguistic Review* 34: 449–77.
Platzack, Christer. 2014. A feature driven account of mainland Scandinavian extraction from relative clauses. In *Talk Given at the Grammar Seminar at the Centre for Languages and Literature*. Lund: Lund University, April 3.
Ross, John R. 1967. Constraints on Variables in Syntax. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.
Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1: 123–34.
Teleman, Ulf, Staffan Hellberg, and Erik Andersson. 1999. *Svenska Akademiens Grammatik [Swedish Academy Grammar]*. Stockholm: Norstedts.
Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.
Thráinsson, Höskuldur, Hjalmar P. Petersen, Jógvan í Lon Jacobsen, and Zakaris Svabo Hansen. 2004. *Faroese. An Overview and Reference Grammar*. Tórshavn: Føroya Fróðskaparfelag.
Van Valin, Robert D. 1994. Extraction restrictions, competing theories and the argument from the poverty of the stimulus. In *The Reality of Linguistic Rules*. Edited by Susan D. Lima, Roberta L. Corrigan and Gregory K. Iverson. Philadelphia: Benjamins, pp. 243–259.

Vincent, Jake W. 2021. Extraction from Relative Clauses: An Experimental Investigation into Variable Island Effects in English-Or-This Is a Dissertation That We Really Needed to Find Someone Who'd Write. Ph.D. thesis, University of California, Santa Cruz, CA, USA.

Vincent, Jake W., Ivy Sichel, and Matthew W. Wagers. 2022. Extraction from English RCs and cross-linguistic similarities in the environments that facilitate extraction. *Languages* 7: 117.

Wellander, Erik. 1939. *Riktig Svenska: En Handledning i Svenska Språkets vård* [*Real Swedish*]. Stockholm: Norstedt & Söner.

Zaenen, Annie. 1985. *Extraction Rules in Icelandic*. New York: Garland.

Zaenen, Annie, Elisabet Engdahl, and Joan M. Maling. 1981. Resumptive pronouns can be syntactically bound. *Linguistic Inquiry* 12: 679–82.

languages

MDPI

Article

Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English

Christiane Müller * and Clara Ulrich Eggers

School of Communication and Culture, Aarhus University, DK-8000 Aarhus, Denmark; cue@cc.au.dk
* Correspondence: christiane.muller@cc.au.dk

Abstract: Adjuncts and relative clauses are traditionally classified as strong islands for extraction across languages. However, the Mainland Scandinavian (MSc.) languages have been reported to differ from e.g., English in allowing extraction from adjunct and relative clauses. In order to investigate the distribution of possible island extractions in these languages based on naturally produced material, we conducted two exploratory corpus studies on adjunct and relative clause extraction in Danish and in English. Results suggest that both extraction from finite adjuncts and from relative clauses appears at a non-trivial rate in naturally produced Danish, which supports the claim that these structures are not strong islands in Danish. In English, we also found a non-trivial amount of examples displaying extraction from finite adjuncts, as well as a small number of cases of relative clause extraction. This finding presents a potential challenge to the claim that English differs from MSc. in never allowing extraction from strong islands. Furthermore, our results show that both languages appear to share certain trends that can be observed in the extraction examples regarding the type of extraction dependency, the type of adjunct clause featured in adjunct clause extraction, and the type of matrix predicate featured in relative clause extraction.

Keywords: adjunct clauses; corpus study; Danish; English; islands; relative clauses

Citation: Müller, Christiane, and Clara Ulrich Eggers. 2022. Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English. *Languages* 7: 125. https://doi.org/10.3390/languages7020125

Academic Editors: Anne Mette Nyvad, Ken Ramshøj Christensen, Juana M. Liceras and Raquel Fernández Fuertes

Received: 8 February 2022
Accepted: 5 May 2022
Published: 18 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Adjunct clauses and relative clauses have traditionally been considered *strong islands* for extraction across languages, based on data such as (1) from English.

1. a. *[Who]$_i$ did Mary cry [after John hit $\__i$]? (Adjunct island) (Huang 1982, p. 503)
   b. *[Which book]$_i$ did John meet a child [who read $\__i$]? (Relative clause island) (Boeckx 2012, p. 5)

In (1a) and (1b), a dependency cannot be established between the initial phrase *Who* (1a) or *Which book* (1b) and its thematic position in the embedded clause ($\__i$). The unacceptability of extraction from adjuncts and relative clauses has traditionally been captured by syntactic island constraints such as the *Adjunct Condition* (Cattell 1976; Huang 1982; Chomsky 1986) and the *Complex NP Constraint* (Ross 1967), banning adjunct and relative clause extraction universally.

However, based on examples like (2), it has been found that Danish and the other Mainland Scandinavian (MSc.) languages allow extraction from both adjunct clauses (2a) and relative clauses (2b) under certain conditions.

2. a. [Den vase]$_i$ får du ballade [hvis du taber $\__i$].
   this vase get you trouble if you drop
   'You are in trouble if you drop this vase.'
   (Danish; Hansen and Heltoft 2011, p. 1814)
   b. [Suppe]$_i$ kender jeg mange [der kan lide $\__i$].
   soup know I many who can like
   'I know many people who like soup.'
   (Danish; Erteschik-Shir 1973, p. 67)

Examples such as (2) appear to violate the Adjunct Condition and the Complex NP Constraint, respectively, yet are intuitively acceptable to native speakers. Prima facie, this seems to suggest that there is cross-linguistic variation in island effects, thus challenging the assumption that island constraints are universal.

Recent investigations on the possibility of island extraction in MSc. have primarily focused on formally and informally collected judgments of constructed sentences. However, even though native speakers agree in describing such structures as acceptable in MSc. and examples are easily found in authentic material (e.g., Jensen 1998, 2002; Lindahl 2017; Kush et al. 2021), formal studies of adjunct and relative clause islands in MSc. commonly yield acceptability ratings that are unexpectedly low (e.g., Poulsen 2008; Christensen and Nyvad 2014; Müller 2015, 2019; Tutunjian et al. 2017; Wiklund et al. 2017; Kush et al. 2018). This apparent mismatch in formal and informal data raises questions about how this contrast can be explained and under which conditions extraction is possible in MSc., and shows that the findings from acceptability studies in this field should be complemented with data obtained using different methods. At the same time, there is also a range of anecdotal evidence of purportedly acceptable extractions from finite adjunct clauses and relative clauses in English (e.g., Chaves and Putnam 2020), raising further questions as to how different English and MSc. really are in this regard.

To further investigate the space of possible island extractions in these languages, we conduct an exploratory corpus study on adjunct clause and relative clause extraction in one of the MSc. languages, Danish, and in English. We expect that an exploration of naturally produced examples of island extractions will provide further insights into what is in fact possible in Danish and English, and how the distribution of possible island extractions can be characterized. To contextualize our corpus study, we first provide a review of recent research on the topic of strong island extraction in the MSc. languages and in English, and identify some remaining open questions with regard to this topic.

2. Background

2.1. Island Extraction in MSc.

Island constructions in the MSc. languages started to attract attention among linguistic researchers in the 1970s and 1980s, when reports accumulated that these languages are unusually permissive with regard to extraction from relative clauses (e.g., Erteschik-Shir 1973, 1982; Maling and Zaenen 1982; Taraldsen 1982; Engdahl 1997) and adjunct clauses (e.g., Hagström 1976; Anward 1982; Jensen 1998, 2002). It is now seen as uncontroversial that extraction from at least some types of adjunct clauses and relative clauses is acceptable in MSc. (although the data on adjunct clause extraction is a bit more sparse than on relative clause extraction). The sentences in (3–4) show some Danish examples reported in the literature.

3. *Relative clause extraction*

a. [Suppe]$_i$ kender jeg mange [der kan lide $\_i$].
soup know I many who can like
'I know many people who like soup.'
(Erteschik-Shir 1973, p. 67)

b. [Det nummer]$_i$ har jeg engang set en artist [der udførte $\_i$].
that number have I once seen an artist that executed
'Once I have seen an artist who performed that piece.'
(Hansen and Heltoft 2011, p. 1815)

c. [Det]$_i$ kender jeg tusindvis af jødiske mænd, [der ville protestere imod $\_i$].
that know I thousands of Jewish men that would protest against
'I know thousands of Jewish men who would protest against this.'
(Hansen and Heltoft 2011, p. 1815)

4. *Adjunct clause extraction*

a. [Den vase]$_i$ får du ballade [hvis du taber $\_i$].
this vase get you trouble if you drop
'You are in trouble if you drop this vase.'
(Hansen and Heltoft 2011, p. 1814)

b. [De sko]$_i$ kan jeg godt vente [mens du reparerer $\_i$].
those shoes can I well wait while you repair
'I can easily wait while you repair those shoes.'
(Hansen and Heltoft 2011, p. 1814)

c. [Det]$_i$ blev han smaddersur [fordi jeg sagde $\_i$].
that became he very angry because I said
'He became very angry because I said that.'
(Jensen 2002, p. 108)

However, in subsequent formal investigations of island extraction in MSc., a more nuanced picture has emerged.

First, even though adjunct and relative clause extractions are intuitively acceptable to speakers of MSc. and occur in spontaneous speech, formal experimental studies commonly report that MSc. participants rate such extractions as less acceptable than one would expect if adjunct and relative clauses were not islands in these languages. Formal judgments for extraction data thus appear inconsistent with the informal judgments reported in the literature. For instance, sentences involving extraction from relative clauses are intuitively acceptable in Swedish (Allwood 1982; Teleman et al. 1999) and Danish (Erteschik-Shir 1973; Nyvad et al. 2017), but scored unexpectedly low ratings in experimental studies by Christensen and Nyvad (2014), Müller (2015), Tutunjian et al. (2017), and Wiklund et al. (2017). A similar situation obtains for adjunct clause extractions, which have been reported to be acceptable, at least for certain types of adjuncts, in Swedish (e.g., Anward 1982; Teleman et al. 1999) and Danish (Jensen 1998, 2002; Nyvad et al. 2017), but yielded acceptability ratings on the lower end of the scale in the formal studies by Müller (2019) on Swedish and by Poulsen (2008) on Danish. In a formal study for Norwegian by Kush et al. (2018), extraction from relative clauses and from conditional adjunct clauses yielded not only acceptability scores that were at the bottom end of the scale, but also superadditive island effects.[1] It is still debated how these mismatching data from formal and informal judgments can be explained.

Second, a few recent studies of the phenomenon indicate that the acceptability of both adjunct and relative clause extraction in MSc. appears to vary as a function of several factors, among others the specific type of adjunct clause (in the case of adjunct clause extraction), the matrix verb (in the case of relative clause extraction), and the type of extraction dependency.

Specifically, in recent studies on Norwegian (Bondevik et al. 2020) and Swedish (Müller 2017), topicalization from conditional adjunct clauses headed by *om* 'if' and from temporal adjuncts headed by *når* 'when' (in Norwegian) or *efter* 'after' (in Swedish) yielded ratings on the upper half of the scale, whereas topicalization from clauses introduced by *fordi* 'because' in Norwegian and *eftersom* 'because' in Swedish received ratings on the low end of the scale. This has given rise to conjectures that different types of adjunct clauses, or adjunct clauses headed by different subordinators, may vary with regard to extraction possibilities (e.g., Bondevik 2018; Bondevik et al. 2020; Müller 2017, 2019).

In the case of relative clause extraction, it has repeatedly been observed that extraction may be more acceptable or more common with certain types of matrix verbs or certain embedding environments (Erteschik-Shir 1982; Engdahl 1997; Hofmeister and Sag 2010; Kush et al. 2013; Löwenadler 2015). Lindahl (2017) investigated relative clause extraction in Swedish based on a collection of naturally occurring examples and found that the major part (around 75%) of her sample of 100 extractions involved extraction from an existential/presentational relative clause, see the example in (5).

5. [honungssenap]$_i$ är det många [som gillar $\_i$ till den där].
honey mustard is there many that like to the there
'There are many people who like honey mustard with that.'
(Lindahl 2017, p. 77)

Existential/presentational relative clauses serve to introduce new referents into the discourse and typically feature an existential verb such as *be* as the matrix predicate introducing the relative clause. Lindahl also found examples of relative clause extraction involving non-existential main verbs in her sample, e.g., with *beundra* 'admire' (6a) or *störa sig på* 'be annoyed by' (6b), but these occurred much less frequently.

6. a. [Det]$_i$ **beundrar** jag folk [som klarar $\__i$ rent psykiskt],
that admire I people that manage purely psychologically
att bara vänta
to just wait
'I admire people who can deal with that psychologically, to just wait.'

b. [det]$_i$ **stör** **jag** **mej** **på** folk [som säger $\__i$]
that annoy I me on people that say
'People who say that annoy me.'

(Lindahl 2017, p. 89)

Finally, there are indications that the acceptability of island extraction in MSc. might also depend on the type of extraction. Studies on Norwegian by Kush et al. (2018, 2019) and Bondevik et al. (2020) suggest that topicalization from adjunct clauses is easier to accept than *wh*-question formation. Lindahl (2017, p. 47) found that in her sample of 100 naturally occurring relative clause extractions in Swedish, topicalization was by far the most common form of extraction (with 93 cases), followed by relativization with 7 instances. None of her examples involved *wh*-extraction from the relative clause, although Lindahl (2017, p. 166) claims that it is possible to construct acceptable examples (see Engdahl 1997 for a similar observation).

Apart from Lindahl (2017), the only other systematic studies of naturally occurring island extractions in MSc. to our knowledge are Lindahl (2010), Jensen (1998, 2002) and Kush et al. (2021). By and large, these studies seem to lend support to the trends previously observed. Lindahl (2010) collected instances of extraction from presentational/existential relative clauses in the Swedish *Parole corpus* at *Språkbanken* and found 134 examples involving topicalization (other dependency forms were not included in the search). Jensen (1998, 2002) has investigated extraction in Danish, based on a material comprised of 9 interviews from the spoken language corpus BySoc, 2 conversations and an excerpt from a TV program. In the 18 h of spoken Danish that her material amounts to, Jensen found 230 instances of extraction from embedded clauses. Only 10 out of these 230 extraction instances involved extraction from a relative clause, and the only case of adjunct clause extraction found involved a non-finite adjunct clause (Jensen 1998, p. 18). Finite adjunct clause extraction is thus unattested in Jensen's material, although Jensen claims that extraction from finite adjunct clauses is acceptable in Danish under certain conditions, based on introspective judgments and examples reported in the traditional literature. While relative clause extraction does appear in her sample, it occurs at a much lower frequency than extraction from declarative complement clauses. Furthermore, all 10 cases of relative clause extraction involved extraction from an existential/presentational relative clause with *være* 'be' as the matrix predicate, see the examples in (7) (glosses our own).

7. a. og [det]$_i$ **var** der sgu nogen [der ikke forstod $\__i$]
and that were there PRT some that not understood
'And there were some who didn't understand that.'

(Jensen 2002, p. 107)

b. [det]$_i$ **er** der jo ingen [der kan vide $\__i$ vel]
that is there PRT nobody that can know PRT
'Nobody can know that, can they.'

(Jensen 1998, p. 22)

c. Men [dem]$_i$ **er** der overhovedet ingen [der snakker med $\__i$]
but them is there at all nobody that talks with
'There is nobody at all who talks to them.'

(Jensen 1998, p. 23)

Kush et al. (2021) recently investigated the occurrence of relative clause extraction (as well as extraction from embedded questions) in Norwegian by searching a child fiction corpus for extraction instances. Relative clause extraction was attested in Kush et al.'s material, however, with 63 examples considerably less frequently than extraction from non-

island declarative complement clauses (411 instances, Kush et al. 2021, p. 17). Their findings moreover showed that relative clause extraction was rather restricted in terms of the type of extraction dependency: In all cases of relative clause extraction found, the extracted element was topicalized. *Wh*-extraction and relativization from relative clauses remained unattested in their sample. In that, relative clauses differed from complement clauses, for which all three types of extraction dependencies occurred at a comparable rate in Kush et al.'s material. Furthermore, the relative clause extractions attested in Kush et al.'s (2021) corpus study also seemed to be very restricted in terms of the embedding environment, as all of the cases of relative clause extraction involved an existential/presentational relative clause (8a) or an *it*-cleft construction (8b).

8. a. [Det]$_i$ er det ingen [som vet $\_i$].
   that is it no one who knows
   'There is no one who knows that.'
   b. [Trist]$_i$ var det bare Ronja [som var $\_i$].
   sad was it only Ronja who was
   'It was only Ronja who was sad.'

(Kush et al. 2021, p. 25)

The corpus studies by Lindahl (2010), Jensen (1998, 2002) and Kush et al. (2021), as well as Lindahl's (2017) collection, show that investigations of naturally produced long-distance dependencies can complement acceptability judgment experiments and provide additional evidence regarding the conditions on relative clause and adjunct clause extraction. However, additional corpus studies are needed to increase our understanding of the phenomenon.

2.2. Island Extraction in English

Both finite adjunct clauses and relative clauses are standardly treated as strong islands in English, banning all extraction (see examples in 1). Truswell (2007, 2011) claims that extraction is acceptable from non-finite adjuncts in English, provided that the matrix and adjunct VP are parts of a single event. According to Truswell, this is the case when the matrix and the adjunct clause event can be interpreted to be related by a *contingent relation* (e.g., causation or enablement), as opposed to a purely temporal relation. However, English disallows any extraction from finite adjuncts according to Truswell, regardless of the semantic relation between matrix and adjunct clause (see also Ernst 2022). For instance, extraction is reported to be acceptable from the non-finite *after*-adjunct clause in (9a) under the premise that John's going home was caused by him talking to someone, but extraction is unacceptable from the finite counterpart of that clause in (9b) (from Truswell 2007, p. 166; see also Manzini 1992).

9. a. [Who]$_i$ did John go home [after talking to $\_i$]?
   b. *[Who]$_i$ did John go home [after he talked to $\_i$]?

At first glance, formal acceptability studies seem to support the picture that finite adjunct clauses are islands in English: *Wh*-extraction from *if*-adjunct clauses received ratings on the lower end of the scale in a study by Sprouse et al. (2012), and similarly, *wh*-extraction from *after*-clauses was rated below the mid-point of the scale in studies by Michel and Goodall (2013) and Müller (2019), even when a causal interpretation of matrix and adjunct clause was available. However, Sprouse et al. (2016) compared *wh*-extraction and relativization out of adjunct clauses in English and found that relativization from *if*-clauses does not result in island effects in terms of the factorial definition of islands (although ratings for both relativization and *wh*-extraction remained relatively low). Nyvad et al. (2022) tested extraction in the form of relativization from *if*-, *when*- and *because*-clauses in English while also providing a supporting context for the extraction stimuli. In this study, extraction from *if*-clauses yielded average ratings on the upper half of the scale and showed no significant difference from *that*-clause extraction. The findings by Sprouse et al. (2016) and Nyvad et al. (2022) seem to suggest that at least for relativization, *if*-adjuncts are not absolute islands in English.

To this date, only few formal acceptability studies of relative clause extraction in English exist. Christensen and Nyvad (2022) investigated *wh*-extraction and topicalization from English relative clauses and found that both types of extraction resulted in ratings on the lower end of the scale, on par with ungrammatical control sentences. Furthermore, the ratings did not differ significantly across different types of transitive matrix verbs. Christensen and Nyvad (2022) take these results to support the standard assumption that relative clauses are strong islands in English. However, Vincent (2021) provides initial evidence that relative clause extraction in English may possibly be ameliorated in the same environments in which also MSc. relative clause extractions have been observed to be particularly felicitous. Vincent (2021) employs the factorial design developed by Sprouse (2007) and Sprouse et al. (2012, 2016) to measure island effects in *wh*-extraction from three different types of relative clauses: relative clauses in existential environments (10a), relative clauses attached to DP predicates in the complement of a copula (10b), and relative clauses embedded under transitive predicates (10c).

10. a. [Which painting]$_i$ do you think that there is only one art collector [who bid on $\__i$]?
    b. [Which painting]$_i$ do you think that Courtney believes that she is the only art collector [who bid on $\__i$]?
    c. [Which painting]$_i$ do you think that Courtney saw the only art collector [who bid on $\__i$]?
    (Vincent 2021, p. 68)

Vincent found statistically significant superadditivity effects (taken to be diagnostic of island effects in the factorial design) for all three tested types of relative clause extraction; however, the size of the effect differed between the three environments. Extraction from existential or predicative relative clauses (type 10a and 10b) yielded smaller island effects than extraction in the transitive environment (type 10c) (Vincent 2021, p. 74). Moreover, extraction from existential or predicative relative clauses yielded higher average ratings than transitive extraction. Vincent takes these results to indicate that at least relative clauses in existential environments may not be strong islands in English.

The picture is further complicated by informal reports of acceptable island extractions in English. Despite the standard assumption that adjuncts and relative clauses are islands in English, there is a range of anecdotal evidence of purportedly acceptable extractions from finite adjunct clauses (11) and relative clauses (12) in English (see also Chaves and Putnam 2020, pp. 67, 91). The examples in (11) and (12) were either attested in authentic language use, or are reported to be acceptable to at least some speakers of English, which raises further questions as to how different English and MSc. really are in this regard.

11. *Adjunct clause extraction*
    a. This is [the watch]$_i$ that I got upset [when I lost $\__i$].
    (Truswell 2011, p. 175, fn.1)
    b. [the details and the whole]$_i$, which an artist cannot be great [unless he reconciles $\__i$]
    (Haegeman 2004, p. 70)
    c. a stranger, from [that remote and barbarian Isle]$_i$ which the Imperial Roman shivered [when he named $\__i$], paused.
    (Haegeman 2004, p. 70)

12. *Relative clause extraction*
    a. Isn't that [the song]$_i$ that Paul and Stevie were the only ones [who wanted to record $\__i$]?
    (Chung and McCloskey 1983, p. 708)
    b. This is [a paper]$_i$ that we really need to find someone [who understands $\__i$].
    (Chung and McCloskey 1983, p. 708)
    c. That's [one trick]$_i$ that I've known a lot of people [who've been taken in by $\__i$].
    (Chung and McCloskey 1983, p. 708)
    d. This is [the child]$_i$ who there is nobody [who is willing to accept $\__i$].
    (Kuno 1976, p. 423)
    e. Then you look at what happens in languages that you know and [languages]$_i$ that you have a friend [who knows $\__i$].
    (McCawley 1981, p. 108)
    f. Violence is [something]$_i$ that there are many Americans [who condone $\__i$].
    (McCawley 1981, p. 108)
    g. This is [the one]$_i$ that Bob Wall was the only person [who hadn't read $\__i$].
    (McCawley 1981, p. 108)
    h. John is [the sort of guy]$_i$ that I don't know a lot of people [who think well of $\__i$].
    (Culicover 1999, p. 230)

In all these examples, extraction is in the form of relativization from the island. This matches the above-mentioned finding from formal studies that island extraction in English, to the extent that it is possible, seems to be restricted to or more available with relativization dependencies (Sprouse et al. 2016; Nyvad et al. 2022).

2.3. Interim Summary

Recent investigations on the possibility of island extraction in MSc. and English have primarily focused on formally and informally collected judgments of constructed sentences, with the exception of the studies by Lindahl (2010, 2017), Jensen (1998, 2002), and Kush et al. (2021) that investigated naturally produced language. Findings show that the traditional picture, according to which island extraction is impossible in English and widely acceptable in MSc., is not as clear-cut as previously thought. In formal studies on islands in MSc., relative clause and adjunct clause extraction constructions were shown to receive very low ratings, or seem to be restricted to specific environments. On the other hand, the island status of adjuncts and relative clauses in English is challenged by anecdotal reports of acceptable cases of extraction and by studies showing that at least some form of adjunct and relative clause island extraction may be acceptable under certain conditions in English. All in all, these data raise new questions regarding what is in fact possible in MSc. and English, and how different these languages really are with regard to island constraints. To address these questions, the findings from acceptability studies should be complemented with data obtained using different methods.

A study of extractions in authentic material may be able to circumvent some of the potential issues associated with investigating the phenomenon by means of formal acceptability studies. For example, it has been suggested that the low ratings obtained for island extractions in some of the acceptability studies on MSc. are due to using stimulus sentences that are overall unnatural or unusually complex (see e.g., Poulsen 2008, p. 96; Müller 2019, p. 69; Tutunjian et al. 2017; Wiklund et al. 2017; Kush et al. 2018). Another factor that has been suggested to explain the low ratings in formal studies is the absence of contextual cues for the stimulus sentences, which may be required for some forms of extraction (e.g., Tutunjian et al. 2017; Wiklund et al. 2017; Kush et al. 2018). By investigating extractions in authentic material, we should be able to get a better picture of what natural extraction sentences look like (in terms of their syntactic and semantic/pragmatic properties), and what kind of contexts they are typically embedded in. Evidence from production data may thus help us to better understand the mismatch often encountered between formal and informal judgments of MSc. island extractions, as well as to develop more natural-sounding stimuli for future formal studies. Another potential issue with formal acceptability studies is that traditional Scandinavian grammars often associate both adjunct and relative clause extraction with colloquial style, or even advise language users to avoid such extractions (see Lindahl 2017; Müller 2019 for an overview). It is thus possible that participants in acceptability experiments let such prescriptive rules influence their ratings, even though speakers may still produce such constructions—especially in informal settings. Corpus studies can make these production data available.

The existing investigations of strong island extractions in naturally produced material (Jensen 1998, 2002; Lindahl 2010, 2017; Kush et al. 2021) show that corpus studies can provide interesting insights in this regard; however—with the exception of Jensen (1998, 2002)—these studies have focused on relative clause extraction while disregarding adjunct islands, and none of them investigate naturally produced island extractions in English.

We conduct two exploratory corpus studies on extraction from strong islands in one of the MSc. languages, Danish, and in English, with the goal to further investigate the distribution of possible island extractions in these languages. In the first study, we investigate adjunct clause extraction in Danish and English corpora. In the second study, we focus on relative clause extraction. We expect that an exploration of naturally produced examples of island extractions will provide further insights into what is in fact possible in Danish and English, and what patterns or trends can potentially be observed among the

found extraction instances regarding the syntactic and semantic/pragmatic properties of such constructions.

3. Corpus Study 1—Adjunct Clause Extraction

3.1. Materials and Methods

The Danish corpus study of adjunct clause extractions was conducted using the written language corpus *KorpusDK* (56 million words) and the spoken language corpora *BySoc* and *SamtaleBank* (MacWhinney and Wagner 2010.)BySoc consists of transcriptions of ca. 80 spoken conversations (1.3 million words) and SamtaleBank of transcriptions of 24 conversations (altogether 6 h, 20 min). The English data were collected using the *British National Corpus* (British National Corpus 2001, 100 million words) and the *Corpus of Contemporary American English* (COCA 2008, COCA, 1 billion words), which both contain text from spoken and written language of various genres. For both languages, the query was complemented by examples retrieved from a Google search.

To search for examples of adjunct island extraction, we employed a combination of different search strings designed to target extraction from finite *if*-, *when*-, and *because*-clauses, and their Danish counterparts, respectively. Our query was restricted by the fact that the corpora we used are not annotated in a way that makes it possible to search for extraction constructions directly. As a consequence of this, we cannot provide quantitative data on the different types of extraction, or directly compare frequencies across constructions or across languages. Rather, our results provide informal insights into what appears to be possible at all or common among the found extraction instances. We start by describing our strategies for searching adjunct clause extractions in Danish, before moving on to our search strategies for the English material.

The search strings used in KorpusDK were mostly based on the example string given in (13), designed to retrieve instances of topicalization from Danish *hvis* ('if')-clauses, and subsequent modifications of it.

13. [ortho="(Den | Det | De)"] []{0, 2} [pos="N"] [pos="V"] [pos="PERS"] []{0,3} [ortho="hvis"]

This search string will return constructions initiated by a noun phrase consisting of one of the Danish determiners *Den*, *Det* or *De*, followed by up to two optional unspecified words (to allow for e.g., potential adjectival modifiers), and a noun. The noun phrase is followed by a verb (since Danish is a verb-second language), a personal pronoun (in order to target sentences with a pronominal subject), up to three optional unspecified words (in order to allow for e.g., potential sentence adverbs or auxiliary verbs in the matrix), and finally by *hvis* 'if'. The search was restricted to constructions with a pronominal matrix subject, since most acceptable examples of adjunct clause extraction reported in the literature also involve pronouns in the matrix subject position. The query was repeated with other possible determiners in the initial position (viz. *Denne/Dette* 'This', *Disse* 'These', *Sådan/Sådanne* 'Such') as well as with the lowercased versions of all determiners mentioned to account for extractions occurring after a conjunction. Moreover, search strings were constructed where the noun phrase construction was replaced by a simple pronoun, i.e., *Mig* 'Me', *Dig* 'You', *Hende* 'Her', *Ham* 'Him', *Den* 'It', *Det* 'It', *Os* 'Us', *Jer* 'You', or *Dem* 'Them', as well as the lower-cased versions of all of these.

In order to search for *wh*-question formation out of adjunct clauses, the search string described above was modified such that the determiner position in the initial noun phrase was replaced by the Danish *wh*-elements *Hvilken*, *Hvilket* and *Hvilke* 'which' (as well as their lowercased counterparts), or such that the entire noun phrase was replaced by the question word *Hvad/hvad* 'what'.

In order to search for relativization from adjunct clauses, the following search string (initiated by the Danish relative complementizer *som*) was employed:

14. [ortho="som"][pos="PERS"] []{0, 3} [pos="V"] []{0, 3} [ortho="hvis"]

Finally, all of the above-mentioned queries were repeated with the complementizers *når* 'when' and *fordi* 'because' replacing *hvis* 'if', in order to also retrieve potential instances of extraction from adjunct clauses introduced by *når* 'when' or *fordi* 'because'. The returned hits were filtered manually for any instances of adjunct clause extraction.

The other two Danish corpora used in this study, BySoc and SamtaleBank, are not grammatically annotated and the search strategy employed for KorpusDK could thus not be applied to them. Instead, we searched these corpora in their entirety for all instances of *hvis*, *når*, and *fordi* in order to find all instances of adjunct clauses introduced by these elements, and then searched the list of resulting hits manually for cases of extraction.

These corpus queries were complemented by a Google search for adjunct clause extractions. The search strings used in these Google queries were for the most part constructed such that they yielded an adjunct clause combined with a matrix that was composed of a pronominal subject and an adjectival psych-predicate of the sort *være/blive glad* 'be/become happy'. Many of the acceptable instances of adjunct clause extraction reported in the literature involve an adjectival psych-predicate of this type, see e.g., the examples in (15).

15. a. [Det]$_i$ vil jeg **blive** **glad** [hvis du vil gøre $_i$].
that will I become happy if you would do
'I will be glad if you do that.'
(Jensen 1998, p. 17)

b. Men [det]$_i$ **bliver** han så **vred**, [naar man siger $_i$].
but that becomes he so angry when one says
'But he gets so angry when one says that.'
(Jensen 1998, p. 17)

c. [Det]$_i$ **er** hun **sur** [fordi du har sagt $_i$].
that is she mad because you have said
'She is mad because you have said that.'
(Jensen 1998, p. 19)

We now turn to the search strategies employed in the English corpora. As the query syntax for COCA and BNC is different from the one for KorpusDK, different search strings had to be constructed to search the English corpora. We also decided to restrict the English search to extraction in the form of relativization and *wh*-extraction, since topicalization is rather marked in English (see e.g., Engdahl 1997; Poole 2017, p. 15), and it was deemed unlikely that the corpora would contain any instances of topicalization from an island. In (16), two examples of the search strings that were used to target *wh*-extraction from *if*-clauses are shown.

16. a. Which NOUN _vd _pp _vv if
b. Which NOUN _vm _pp _vv if

These strings target question constructions with a *Which*-NP and a form of *do* (_vd) or a modal verb (_vm) as the finite matrix verb and a personal pronoun (_pp) as the matrix subject. These search strings were subsequently augmented with additional unspecified positions in the matrix predicate.

Relativizations from adjunct clauses were targeted with strings like the following, involving a noun followed by one of the relative complementizers *that*, *which* or *who*.

17. NOUN that | which | who _pp _vv if

Like with the Danish part of the study, the same search strings were also employed with the alternative adjunct clause subordinators *when* and *because*.

As in the Danish study, Google was used to find further examples of adjunct clause extractions. The search strings used for Google mostly targeted relativization with *which* and a matrix clause involving a pronominal subject and, parallel to the Danish Google search, a psych-predicate such as e.g., *be glad*.

3.2. Results

3.2.1. Adjunct Clause Extraction in Danish

Our search on KorpusDK, BySoc, and SamtaleBank combined with the Google search yielded in total 36 instances of sentences displaying extraction from adjunct clauses in Danish. Out of these, only 3 instances were retrieved from KorpusDK, 1 from BySoc, and the remaining 33 examples were retrieved from the Google search. No instances were found on SamtaleBank.

In 7 instances, the extracted phrase is an adjunct rather than an argument in the embedded adjunct clause, see the examples in (18). The sentence-initial adjunct in both (18a) and (18b) must be interpreted as modifying the predicate in the adjunct clause, rather than the matrix predicate.

18. a. [Til vores fredagsarrangementer, hvor vi henter mange kendte
to our Friday events, where we get many known
foredragsholdere]$_i$, skal vi være glade, [hvis der kommer fem,
speakers shall we be glad if there come five
seks herrer _ $_i$].
sex men
'We should be glad if five, six men come to our Friday events, where we get many well-known speakers.'
[https://sn.dk/Rudersdal-Avis/De-giver-livet-indhold-Andres-saavel-som-deres-eget/artikel/1181760, accessed on 3 December 2020]

b. [Der]$_i$ er vi meget glade, [hvis vi får noget publiceret _ $_i$].
there are we very happy if we get something published
'We are very happy if we get something published there.'
[http://curis.ku.dk/ws/files/59251974/JensPeterAndersenThesis.pdf, accessed on 3 December 2020]

Regarding the adjunct clause type that is featured in the extraction sentences, 31 sentences involved extraction from *hvis* ('if')-clauses and 6 sentences featured *når* ('when')-adjunct clauses. No cases of extraction from *fordi* ('because')-clauses were found. As for the extraction dependency, 16 instances involved topicalization and 21 involved relativization. None of the extractions were in the form of *wh*-extraction. Table 1 shows the distribution of the different clause types and extraction dependencies in the sample.

Table 1. Distribution of the Danish adjunct clause extractions.

	***hvis* 'If'**	***når* 'When'**	***fordi* 'Because'**	**Total**
wh-extraction	0	0	0	0
topicalization	13	3	0	16
relativization	18	3	0	21
total	31	6	0	37

Sentences (19a–d) demonstrate an example from each category that was attested.

19. a. Topicalization from *if*-clause
[især den sidste]$_i$ ville jeg blive sur, [hvis du ikke får _$_i$].
especially the last will I get mad if you not get
'I will get mad if you do not get the last one in particular.'
[http://www.thevintagehausfrau.dk/2010/09/23/gaveoensker/, accessed on 15 November 2020]

b. Relativization from *if*-clause
Det er [et nødvendigt mineral]$_i$, som vi ville dø, [hvis vi ikke fik _$_i$].
it is a necessary mineral that we would die if we not got
'It is a necessary mineral that we would die if we did not get.'
[https://www.hvadhvemhvor.dk/hvornaar-maa-boern-faa-salt/, accessed on 30 November 2020]

c. Topicalization from *when*-clause
[det]$_i$ bliver jeg glad [når jeg læser _$_i$] for sådan har jeg det også
this become I happy when I read because like that have I it too
'I get happy when I read this because I feel the same way.'
[http://tumle.bloggersdelight.dk/jeg-bliver-edderspaendt-rasende-og-en-smule-ked-af-det-eller-omvendt/, accessed on 15 November 2020]

d. Relativization from *when*-clause
Ledelsen på CC har FLERE gange offentligt udtalt, at tronc systemet
board.the at CC has several times publicly announced that tronc system.the
er [en guideline]$_i$, som de bliver glade [når spillerne følger _$_i$],
is a guideline that they become happy when players.the follow
men IKKE et krav.
but not a requirement
'The board at CC has several times publicly announced that the tronc system is a guideline that they are happy when the players follow, but not a requirement.'
[https://www.pokernet.dk/forum/smidt-ud-af-casinoet.html, accessed on 3 December 2020]

3.2.2. Adjunct Clause Extraction in English

In total, 49 sentences involving extraction from an adjunct clauses where found in English, out of which 5 stem from COCA, 1 from BNC and the remaining 43 from Google. A caveat with using Google to retrieve English examples is of course that it cannot be guaranteed that all examples were authored by native English speakers (whereas this is unlikely to be an issue for Danish). We tried to minimize the risk of including examples produced by non-native speakers of English by examining the source, the context, and (where possible) the author of all English examples, and consequently excluding any cases where we had reasons to suspect that they were authored by a non-native speaker. However, we concede that it is not possible to fully rule out the possibility of including examples from non-native speakers in our sample, and the results retrieved from Google for the English part of this study thus allow only for limited conclusions.

Regarding the adjunct clause type that is featured in the extraction sentences, 42 sentences involved extraction from *if*-clauses and 7 sentences featured *when*-adjunct clauses. No cases of extraction from *because*-clauses were found. As for the extraction dependency, all English cases involved relativization, and no examples of *wh*-extraction from adjunct clauses were found (as mentioned above, we did not search for instances of topicalization in English). Table 2 shows the distribution of the different clause types and extraction dependencies in the sample.

Table 2. Distribution of the English adjunct clause extractions.

	If	***When***	***Because***	**Total**
wh-extraction	0	0	0	0
relativization	42	7	0	49
total	42	7	0	49

Sentences (20a–b) demonstrate an example of the two types of adjunct clause extraction found in our sample.

20. a. Relativization from *if*-clause
Many of the exercises are [ones]$_i$ that I would be surprised [if even 1 percent of healthy women can do $\_i$].
[COCA 2008]

b. Relativization from *when*-clause
Now, those are [things]$_i$ that I feel very warm [when I look at $\_i$], and I wouldn't want to live in a house that they—a house that didn't have room for those.
[COCA 2008]

3.3. Discussion

The fact that we found a variety of adjunct clause extraction sentences in Danish demonstrates that adjunct island extraction in MSc. is not just a peripheral phenomenon restricted to isolated constructed examples, but that naturally occurring cases can be attested in authentic language use. To have a rough point of comparison for the frequency of occurrence of the Danish island extractions, another search was done on KorpusDK that targeted extraction from (non-island) declarative complement clauses introduced by *at* 'that'. To this end, the search strings earlier employed for adjunct clause extractions were reused, but with *at* 'that' replacing the adjunct clause subordinator *hvis* 'if', *når* 'when' or *fordi* 'because'. Note that this excludes any instances of extraction from complement clauses with a non-overt *at* 'that' from the search results. This search resulted in 1250 instances of extraction from declarative *at*-clauses in KorpusDK. Not taking into account the base frequency of complement clauses vs. adjunct clauses in Danish, it can thus be asserted that extraction from (some) adjunct clauses appears to be possible in Danish, but occurs considerably less frequently than *at*-clause extraction.

Perhaps more surprisingly, a non-trivial amount of naturally produced examples of adjunct clause extraction were also found in English. This finding potentially challenges the claim that English differs from the MSc. languages in never allowing extraction from finite adjunct clauses, as argued e.g., by Truswell (2007, 2011) and Ernst (2022). However, in light of the small number of English examples found in relation to the size of the English materials that we used, and the fact that it cannot be guaranteed that all English examples were produced by native speakers, our results do not permit a clear conclusion as to whether extraction from finite adjunct clauses is generally acceptable in English. At the same time, our results are in line with recent experimental findings that at least *if*-adjunct clauses in English do not behave like absolute islands for extraction in the form of relativization (Sprouse et al. 2016; Nyvad et al. 2022).

In both languages, by far most of the extraction examples were found on Google rather than in the investigated corpora. This could in part be due to the corpora not being large enough to feature many examples of a construction as infrequent as adjunct clause extraction. It is also possible that adjunct clause extractions are more common in very informal text types that resemble spoken language (such as blogs and discussion forums), which are more accessible via Google than a corpus consisting mostly of written language.

Some trends and patterns can be discerned among the found extraction instances regarding the type of adjunct clause featured as well as the type of extraction dependency. When it comes to clause type, all examples of adjunct clause extraction found in English as well as in Danish featured *if*- and *when*-clauses, with extraction from *if*-clauses clearly being the most common in both languages (see Tables 1 and 2). The fact that no examples of extraction from *because*-clauses were found in either language is consistent with previous observations that conditional and (certain) temporal adjuncts appear to be more permissive for extraction than causal adjuncts, and that extraction possibilities thus may differ across different types of adjunct clauses (e.g., Müller 2019; Bondevik et al. 2020).

If Danish and English indeed allow extraction from *if*- and *when*-adjunct clauses, at least under some conditions, then this possibility could potentially be accommodated under accounts proposing that adjunct clause extraction is facilitated when a causal relation between the events described in the matrix and in the adjunct clause can be established (e.g., Jensen 1998, 2002; Truswell 2007, 2011). Both conditional *if*-clauses and *when*-clauses (at least *when*-clauses allowing for a generic reading) usually specify general causes or circumstances for a state of affairs expressed in the matrix clause and can thus easily be construed as being causally related to the matrix. A look at the extraction examples that we retrieved from our corpus search confirms that in all cases, a causal interpretation of the events is very natural. Often, the matrix predicate expresses a psychological condition such as *be glad* (see examples 18a–b and 19c–d) or *be angry* (see 19a) that is presented as a consequence of the proposition expressed in the adjunct clause. However, contra Truswell (2007, 2011) and Ernst (2022), extraction from causally interpreted adjuncts in English does not seem to be restricted to non-finite adjuncts.

At first glance, an account of permissible adjunct island extractions in terms of a causation relationship between adjunct and matrix clause leaves unexplained the absence of extraction from *because*-clauses in both English and Danish in our study. Since *because*-clauses explicitly mark a causal relation, they should permit extraction as easily as e.g., *if*-clauses according to the above proposal. However, there are some indications that causal adjunct clauses differ from conditional and temporal adjuncts in possessing a more elaborate internal structure that could be responsible for blocking extraction, see Müller (2017, 2019). For instance, Johnston (1994) and Sawada and Larson (2004) argue that causal clauses differ from e.g., temporal clauses semantically in having a "closed event" structure and asserting, rather than presupposing the existence of the event described in them. As Müller (2017) points out, a closed event structure could possibly also entail syntactic opacity, if semantic closure interacts with cyclic Spell-Out such that it induces Transfer of the relevant structure to the interfaces, thereby making it unavailable for subsequent extraction. Furthermore, Sawada and Larson (2004) suggest that this semantic difference may also

have a syntactic parallel, in that causal connectives such as *because* combine not only with a larger semantic domain than temporal connectives, but also with a larger syntactic domain that contains additional layers of structure. This extended syntactic domain could constrain extraction possibilities, for instance, if it contains a feature or projection that causes an intervention effect with the extracted phrase.

As for the type of extraction dependency, most cases of adjunct clause extraction found involved relativization from the island, followed by topicalization in Danish. Notice that in English, only relativization and *wh*-extraction were included in the search. No cases of *wh*-extraction were found in either language. This finding lends further support to previous observations that extraction possibilities may differ across extraction dependency, with topicalization (Kush et al. 2018, 2019; Bondevik et al. 2020) and relativization (Sprouse et al. 2016; Abeillé et al. 2020) from islands reported to be more acceptable than *wh*-extraction. We return to a discussion of the role of extraction dependency in Section 5.

4. Corpus Study 2—Relative Clause Extraction

4.1. Materials and Methods

The Danish data for relative clause extractions were collected using KorpusDK. In light of the large quantity of relative clause extraction sentences that we found on KorpusDK, we decided to restrict the search to KorpusDK and did not extend the investigation to include BySoc, SamtaleBank or Google for the purpose of finding Danish relative clause extractions. Note also that Jensen (1998, 2002) has previously searched BySoc for extraction instances and retrieved the cases of relative clause extraction occurring in BySoc (see Section 2.1). The English part of the search was once again carried out using the corpora BNC and COCA, as well as Google. Because we were not able to find any instances of relative clause extraction on BNC or COCA, we extended the search to also include the corpus of Global Web-based English (Davies 2013, GloWbE, 1.9 billion words), which contains material from various websites in 20 different English-speaking countries.

For the search on KorpusDK, the search strings described in Section 3.1 were reused with some adaptions to target relative clause extraction instead of adjunct clause extraction. Thus, the position specifying the adjunct clause subordinator (e.g., *hvis*) was replaced by the relative pronouns *som* and *der* and preceded by 1–4 optional words (rather than 0–3) to allow for a head noun of the relative clause, see the example search string in (21) (targeting topicalization from a relative clause).

21. [ortho = "(Den | Det | De)"] []{0, 2} [pos = "N"] [pos = "V"] [pos = "PERS"] []{1, 4}[ortho = "(som | der)"]

Also for English, the search strings used to target adjunct clause extractions (described in Section 3.1) were reused, to the extent that this was possible, and adapted to target relative clauses instead of adjunct clauses. For instance, relativization from relative clauses was targeted with strings like (22a–b), searching for a noun phrase followed by one of the relative complementizers *that*, *which* or *who*. Instead of an adjunct clause subordinator like *if*, the search string ends on a relative clause complementizer (*who* or *that*), which in turn is preceded by a pronoun position (_p), since the head noun in the relative clause extractions reported in the literature is often an indefinite pronoun of some kind.

22. a. NOUN that | which | who _pp _vv _p who
    b. NOUN that | which | who _pp _vv _p that

Like in the adjunct clause search, these strings were augmented with additional unspecified positions in the matrix clause in subsequent queries. Additional searches were carried out targeting specific matrix constructions that appear to be common in relative clause extraction sentences, e.g., by using strings that target relative clauses headed by NPs with *the only* (23a) or *a lot of* (23b), and modifications of these strings.

23. a. that | which | who _pp [be] the only * who | that
    b. that | which | who there [be] a lot of * who | that

Again, Google was used to carry out further searches, mostly targeting relativization from relative clauses in constructions that appear to be common in the extraction sentences found in the literature or in our Danish material, e.g., *which I don't know anyone who* or *which there are many people who*. Like with the adjunct clause study, the English search was restricted to *wh*-extraction and relativization and did not include topicalization.

4.2. Results

4.2.1. Relative Clause Extraction in Danish

In total, we found 940 instances of Danish relative clause extraction on KorpusDK, out of which 910 involved topicalization from the relative clause and 30 involved relativization. Again, no instances of extraction by *wh*-question formation were found. Table 3 shows the distribution of Danish relative clause extractions across the different extraction dependencies and the type of matrix verb under which the relative clause was embedded.

Table 3. Distribution of the Danish relative clause extractions.

	Matrix Verb *være* 'Be'	**Other Matrix Verb**	**Total**
wh-extraction	0	0	0
topicalization	905	5	910
relativization	28	2	30
total	933	7	940

As Table 3 illustrates, the overwhelming majority of extraction instances (933 out of 940) has a form of *være* 'be' as the matrix predicate. In most of these (viz. in 866 cases), the matrix predicate *være* is part of an existential construction with *der* 'there' as the subject, as exemplified in (24).

24. a. Topicalization from a relative clause
[Det]$_i$ **er** **der** allerede mange, [der gør _$_i$].
this are there already many who do
'There are already many who do this.'
[KorpusDK 2021]

b. Relativization from a relative clause
Der var [noget damegymnastik på skærmen]$_i$, som **der** ikke **var** nogen, [der så på _$_i$]
there was some women's gymnastics on screen.the that there not was anybody who looked at
'There was some women's gymnastics on the screen that there wasn't anybody who was looking at.'
[KorpusDK 2021]

In most of the remaining cases with matrix verb *be*, the relative clause is attached to a DP in a copular construction with a referential pronominal subject, see e.g., (25).

25. Men [det]$_i$ er jeg den eneste [der ved _$_i$].
but that am I the only who knows
'But I am the only one who knows that.'
[KorpusDK 2021]

Only 7 cases of extraction were found that involved a different matrix verb than *være* 'be'. The matrix verbs in these were *kende* 'know', *møde* 'meet', *have* 'have', *blive* 'become', and *finde* 'find', see the examples in (26).

26 a. Men [den slags spekulationer]$_i$ **kender** jeg ikke nogen, [der tør
but this sort speculations know I not anybody who dares
fremsætte $\__i$ offentligt].
to.put forward publicly
'But I don't know anybody who dares to put forward these sorts of speculations publicly.'
[KorpusDK 2021]

b. [Det]$_i$ **mødte** jeg nu slet ikke nogen, [der gjorde $\__i$].
that met I now at all not anyone who did
'I did not meet anyone at all who did that.'
[KorpusDK 2021]

c. Og [det]$_i$ **har** vi indvandrere, [der gør $\__i$ hver eneste aften]
and that have we immigrants who do every single evening
'And we have immigrants who do this every single evening.'
[KorpusDK 2021]

d. og [det]$_i$ **bliver** vi flere og flere [der foretrækker $\__i$]
and that become we more and more who prefer
'And there are more and more of us who prefer that.'
[KorpusDK 2021]

e. Det viste sig at ville blive [en meget kostbar affære]$_i$, som vi ikke
it showed REFL to will become a very costly affair that we not
kunne **finde** nogen, [der ville påtage sig $\__i$] .
could find anyone who wanted to.take on REFL
'It turned out to be a very costly affair that we could not find anyone who wanted to take on.'
[KorpusDK 2021]

There was also a clear trend in regard to the extracted element: In 797 cases (ca. 85%), the fronted element was a simple demonstrative or personal pronoun of the type *den*/*det*/*dem* 'that'/'that'/'those' (most commonly, *det*), rather than e.g., a full noun phrase.

Finally, our query also returned 27 sentences in which a manner adverbial adjunct (*sådan* 'like that') (27) or a PP adjunct (28) rather than an argument has been extracted from the relative clause.

27. a. [Sådan]$_i$ er der- endnu- ikke så mange, [der tænker $\__i$]
like that are there yet not so many who think
'There are not yet that many who think like that.'
[KorpusDK 2021]

b. men [sådan]$_i$ er der jo ingen mennesker [der lever $\__i$ i dag].
but like that are there PRT no people who live today
'But there aren't any people who live like that today.'
[KorpusDK 2021]

28. [Mod nerverne]$_i$ er der også noget naturmedicin, [som du kan tage $\__i$].
against nerves.the is there also some natural medicine that you can take
'There is also some natural medicine that you can take against the nerves.'
[KorpusDK 2021]

4.2.2. Relative Clause Extraction in English

Only 18 cases of relative clause extraction were found in the English material, all of which involved extraction by relativization. Almost all of the English examples were found on Google; the search on BNC, COCA and GloWbE only returned one instances of relative clause extraction (found on GloWbE). Only 7 of the English examples involved a form of *be* as the matrix verb (see Table 4). The remaining sentences had a form of *know* (10 cases) or *meet* (1 case) as the matrix predicate.

Table 4. Distribution of the English relative clause extractions.

	Matrix Verb *be*	**Other Matrix Verb**	**Total**
wh-extraction	0	0	0
relativization	7	11	18
total	7	11	18

In all 7 cases with matrix verb *be*, the relative clause modifies a *the only* DP in a copular construction, see the examples in (29).

29. a. And I always make [a plum pudding with hard sauce]$_i$, which **I am the only** person [who eats _ $_i$].
[https://www.ihavenet.com/recipes/Peppermint-Pie-Christmas-Dessert-Recipe-One-for-the-Table-Recipes.html, accessed on 17 November 2021)
b. The compressed-encrypted stream would be as if we are using [a different language (still Zeros and Ones = 0101)]$_i$ which **we are the only** ones [who can decompress-decrypt _ $_i$].
[https://www.linkedin.com/pulse/recent-russian-hacking-our-country-number-us-sam-eldin?trk=public_profile_article_view, accessed on 17 November 2021]

The remaining cases (featuring *know* or *meet* as matrix verb) all involve extraction of a VP from a relative clause headed by *anyone* or *anybody*, as shown in (30).

30. a. I've [done five records in the last six years]$_i$, which I don't **know** anybody [who has _$_i$].
[https://www.spin.com/2021/07/david-crosby-5-albums-i-cant-live-without/, accessed on 19 November 2021]
b. If you [send 20 letters a month]$_i$ (which I don't **know** ANYONE [who does _$_i$]), that would be an extra dollar or two.
[http://voices.washingtonpost.com/federal-eye/2009/05/postal_service_posts_19_b_quar.html, accessed on 19 November 2021]
c. If you [take perfect care of a string]$_i$ (which I've never **met** anybody [who does _$_i$]) then yes it will last forever
[https://www.tapatalk.com/groups/cellofun/strings-for-a-goffriller-t9970.html, accessed on 19 November 2021]

4.3. Discussion

The quantity of sentences featuring relative clause extraction that we found in Danish (940 cases on KorpusDK alone) allows us to conclude that relative clause extraction, especially with topicalization, is a commonly produced construction in naturally occurring Danish. In fact, the frequency of relative clause extraction is roughly comparable to the rate of extraction from declarative *at* ('that')-clauses that we found using equivalent search strings, with 1250 found cases on Korpus DK (see Section 3.3). Note that this does not take into account the base frequency of complement clauses vs. relative clauses in Danish.

The considerable amount of relative clause extractions found on KorpusDK seems to support the previous observations that relative clauses do not behave like strong islands in Danish. However, it is striking that a vast majority of the found relative clause extractions (more than 92%) feature a relative clause embedded under an existential construction introduced by *der er* 'there is'. Relative clause extraction in Danish thus seems to be particularly productive in this specific environment. This is in line with similar observations for Swedish, where relative clause extraction is also reported to be most common in existential environments (e.g., Engdahl 1997; Lindahl 2017). In light of the few extraction examples that we found featuring other matrix verbs (e.g., *kende* 'know', *møde* 'meet', or *finde* 'find'), our data are compatible with the claims that MSc. relative clause extraction is in principle also possible with other (non-existential) predicates (see also Lindahl 2017). However, the production of such cases in written language appears to be exceedingly rare (given that less than 1% in our sample featured a matrix verb other than *være* 'be').

One possible explanation for the uneven distribution of matrix verbs is that there is a syntactic difference between relative clauses embedded under an existential construction and other relative clauses, such that the latter form islands for extraction and the former do not. McCawley (1981) makes a proposal along these lines, by suggesting that extraction from relative clauses is possible when the relative clause is embedded in an existential or negative existential clause, as the extraction domain in that case is not a regular relative clause, but a *pseudo-relative*. Extraction from a pseudo-relative may not violate an island constraint if it is assumed that pseudo-relatives are not actually complex NPs (e.g., Casalicchio 2016).[2] However, this approach would not cover the cases where the matrix verb is a not an existential verb, such as e.g., *finde* 'find', see example (26e). See also Lindahl (2017)

for a review of Swedish extraction examples that cannot be covered by McCawley's (1981) proposal. In other words, the preference for an existential matrix verb in MSc. relative clause extractions seems to be more of a strong tendency than an absolute restriction.

This tendency can more likely be explained by a pragmatic account of islands, as suggested by Chaves and Putnam (2020). Chaves and Putnam propose that many island constraints traditionally assumed to be syntactic in nature, including the Complex NP Constraint responsible for relative clause islands, can be reduced to *Relevance Islands*: The referent that is singled out by the extraction must be sufficiently relevant for the main action described by the utterance. Chaves and Putnam's proposal builds on a line of other accounts that derive island effects from information-structural factors (e.g., Erteschik-Shir 1973; Deane 1991; Goldberg 2006, 2013; Van Valin 1994, 1996, 2005). Generally, these accounts share the assumption that extraction is only felicitous if it occurs from a constituent that is in some sense prominent or relevant in the discourse. For example, Goldberg (2006) suggests that extraction is illicit from *backgrounded* domains, since extracted phrases are typically in discourse-prominent positions, and extraction from a backgrounded domain thus causes a pragmatic clash. Chaves and Putnam (2020) recast the account by Goldberg (2006, 2013) and other related pragmatic accounts of islands in terms of the concept of *relevance*: An extracted referent "must be highly relevant (e.g., part of the evoked conventionalized world knowledge) relative to the main action that the sentence describes" (Chaves and Putnam 2020, p. 206). According to Chaves and Putnam (2020, p. 68), this relevance constraint can account for the difficulty to extract from most relative clauses: Because relative clauses tend to express presupposed or backgrounded information, a referent belonging to a relative clause can typically not be construed as sufficiently relevant for the main event. However, extraction may be acceptable if the relative clause is embedded under an existential *there is/are* or other matrix predicates that are low in semantic content, such as e.g., *know* or *have*, since in those cases, the embedded clause can be deemed to be more informative than the matrix clause, and drawing attention to a referent from it by extraction thus does not violate the pragmatic relevance principle (Chaves and Putnam 2020, p. 68). Indeed, this explanation seems to accommodate not just the cases of relative clause extraction under *there is/are*, but also the examples involving other matrix predicates that we found: All other matrix verbs attested in our extraction examples (*kende* 'know', *møde* 'meet', *have* 'have', *blive* 'become', and *finde* 'find') are semantically rather abstract and are thus compatible with the embedded relative clause expressing the main assertion in the utterance.

As a way to identify relevant or prominent constituents, Chaves and Putnam (2020) and Goldberg (2006, 2013) both suggest that the distinction between discourse-prominent (relevant) and backgrounded content aligns with the distinction between asserted and presupposed content, such that asserted information tends to correspond to the main action (or in Goldberg's terms, the *potential focus domain* of a sentence), whereas presupposed clauses are backgrounded (Chaves and Putnam 2020, pp. 71, 208; Goldberg 2006, p. 130). Assertions in turn can be identified by testing whether a proposition can be negated by sentential negation. This negation test correctly predicts that existential/presentational relative clauses as well as predicative relative clauses (which together represent the bulk of relative clauses involved in our examples) should allow extraction, as they express assertions and can thus successfully be negated by negating the matrix (see Kush et al. 2021, p. 160, Kush et al. 2021, p. 38). However, as Lindahl (2017, pp. 160–61) points out, the negation test runs into difficulties with cleft structures, which appear to allow extraction in the MSc. languages, but are incorrectly identified as islands by the negation test, since they are presupposed (see also Kush et al. 2021, p. 38). While extraction from cleft relative clauses was rare in our material, we did find 8 instances of extraction from a cleft clause in the Danish corpus that would be left unaccounted for under the proposal that islands correspond to presupposed clauses. Consider e.g., example (31a), which involves extraction from a cleft clause. As (31b) shows, it is not possible to negate the proposition expressed in the cleft relative clause by negating the matrix of the non-extracted version of this sentence.

31. a. [Dem]$_i$ er det primært centrum-venstre-partierne, [der har svaret på $\_\__i$].
them is it primarily center-left parties.the that have replied to
'It is primarily the center-left parties that have answered them.'
[KorpusDK 2021]
b. Det er ikke primært centrum-venstre-partierne, [der har svaret på dem].
it is not primarily center-left-parties.the that have answered to them
'It is not primarily the center-left parties that have answered them.'
→ Someone has replied to them.

However, Kush et al. (2021) point out that another related information-structural factor seems to make the right cut between the types of relative clauses that allow for extraction and those that do not, viz. whether or not the clause in question conveys new information: "The [relative clauses] that allow movement are those that contribute wholly or partially new information to the discourse (or at least information that need not be known to the hearer)" (Kush et al. 2021, p. 39). This proposal can account for the possibility to extract from cleft structures, since cleft clauses (despite being presupposed) may convey new information. For example, it is possible to utter the sentence in (31a) even if the information provided in the cleft clause (that someone has answered them) is new in the discourse. Kush et al. (2021) further suggest that clauses can be said to provide new information when they contain the sentence's *main point of utterance* (Simons 2007). Since the matrix predicate in (31a) and in other cleft sentences is almost void of semantic content, the embedded relative clause necessarily constitutes the MPU in cleft sentences like (31a). An account of transparent relative clauses in terms of new information or MPU can arguably also account for our extraction examples where the matrix predicate is not *være* 'be'. As pointed out above, the other matrix verbs found in our extraction examples (e.g., *kende* 'know', *have* 'have', and *blive* 'become') are also low in semantic content and are thus compatible with the relative clause constituting the MPU. Finally, this proposal also has the potential to account for some of the Swedish examples of relative clause extraction that Lindahl (2017) has shown to be problematic for an account in terms of backgrounded or presupposed constituents, viz. examples with *beundra* 'admire' or *störa sig på* 'be annoyed by' as matrix predicates (see 6a–b). As Lindahl (2017, p. 161) shows, the clauses embedded under these verbs fail the negation test and are therefore considered to be backgrounded, yet they seem to allow extraction. However, it seems possible that clauses in the complement of these verbs still may constitute a sentence's MPU: According to Simons (2007), an embedded clause can be the MPU if the matrix predicate conveys the speaker's emotional orientation towards the information in the embedded clause, as the matrix verb is considered parenthetical in that case. This seems to fit the function of the matrix verbs in the Swedish examples mentioned above. Even though these examples are quite different from the ones discussed in Simons (2007), the predicates 'admiring' and 'being annoyed' can also be argued to express an emotional orientation towards the content of the relative clause.

Although we acknowledge that these remarks are of a very preliminary nature and require further investigation, we tentatively suggest (along with Kush et al. 2021) that a pragmatic account in terms of new information or MPU could make the relevant distinction between relative clauses that allow for extraction and those that do not (to the extent that a language allows extraction from relative clauses at all). However, more work is required to flesh out how exactly Simons (2007) proposal could be adapted to relative clauses.

Unlike with adjunct clause extraction, the extraction dependency that was by far most frequent in our sample of Danish relative clause extractions was topicalization, with relativization taking a second place. However, the results from our relative clause study share with the adjunct clause study that none of the found extraction instances involved *wh*-extraction. This finding thus further strengthens the conjecture that topicalization and relativization from islands appears to be easier than *wh*-extraction. Our results regarding the distribution of different extraction dependencies in relative clause extraction match Lindahl's (2017, p. 47) finding that topicalization was by far the most common extraction dependency among her sample of Swedish relative clause extractions, whereas relativization occurred only in a few cases and *wh*-extraction from relative clauses remained unattested.

Finally, our finding that not just arguments, but also adverbial or PP adjuncts are extracted from relative clauses in our sample (see examples 27 and 28) demonstrates that relative clause extraction in Danish apparently is not restricted to argument DPs, but that phrases of different categories can be topicalized (see Lindahl 2017 for a similar observation for Swedish relative clauses). If MSc. relative clauses indeed allow for extraction of (some) adjuncts, as these observations suggest, this would distinguish MSc. relative clauses not just from strong islands, but also from traditional weak islands that generally permit extraction of arguments, but not adjuncts (Huang 1982; Szabolcsi 2006).

As for English, the extraction instances that we found on Google and GloWbE seem to support the anecdotal evidence that naturally produced cases of extraction from English relative clauses do occasionally appear in authentic language use. However, the low frequency of the results prevents a clear conclusion as to whether relative clause extraction in English is a phenomenon that extends beyond constructed examples and only sporadically found natural cases. It should, however, also be mentioned that the frequency for relative clause extractions cannot be directly compared between our Danish and English study, since the search methods we could employ for the English corpus study were somewhat restricted in comparison to the Danish study—in part because of limits on the length and generality of the search strings that can be used to query BNC, COCA and GloWbE, and in part due to the polysemy of English *that*. We were thus not able to conduct a search for English relative clause extractions that was as systematic and extensive as the Danish part of this study.

It is, however, remarkable that no cases of extraction from relative clauses in existential environments were found in English, given that a large part of our search methods specifically targeted such constructions, and that extraction from existential relative clauses was so common in the Danish material. Instead, the majority of the examples we found involved VP-extraction from a relative clause embedded under *know* and headed by *anyone* or *anybody*. In light of the low number of cases that we found in English, we refrain from any further analysis and conclusions regarding potential patterns or trends in English relative clause extraction.

5. General Discussion

Extraction from relative clauses and adjunct clauses is intuitively acceptable to native speakers of the MSc. languages, at least under certain conditions. Such cases of acceptable adjunct and relative clause extraction pose a puzzle for syntactic research, as these structures seem to violate island constraints assumed to apply universally. However, a few recent experimental studies report very low acceptability ratings for island extraction sentences in MSc., which has raised the question how these experimental findings can be reconciled with the observation that adjuncts and relative clauses do not seem to behave like islands in MSc. The finding from our corpus studies that both extraction from finite adjuncts and from relative clauses can be attested at a non-trivial rate in naturally produced Danish lends strong support to the notion that extraction from adjunct and relative clauses, at least in some environments, appears to be possible in Danish, and that these constructions thus are not absolute islands in Danish.

Perhaps more surprisingly, we were able to show that extraction from some types of finite adjunct clauses and relative clauses also occurs in naturally produced English (although the number of both types of extractions we could find in English remained very small). Due to the small number of examples we could find and the fact that most of our English examples were retrieved from Google, our findings do not permit us to draw any further conclusions regarding whether or not English generally allows extraction from traditional strong islands. However, we want to point out that the instances of extraction from finite *if*-adjunct clauses that we found in English are in support of recent experimental findings suggesting that at least *if*-adjunct clauses in English may be transparent for extraction by relativization (Sprouse et al. 2016; Nyvad et al. 2022).

Moreover, many of the trends and patterns that can be observed among our found extraction instances are shared by English and Danish. First, to the extent that extraction from adjunct clauses and from relative clauses is attested in these languages, it seems to occur only with relativization (and in Danish, topicalization) from the island, but remains unattested with *wh*-extraction. Second, adjunct clause extraction in both languages appeared most frequently from *if*-clauses and only in a few instances from a *when*-clause. Extraction from *because*-clauses was unattested in both Danish and English. Third, relative clause extraction occurred only with a very limited number of different verbs as the matrix predicate in both languages. These similarities can be taken as an indication that English and Danish are possibly more similar with regard to strong islands than previously assumed, in the sense that these languages facilitate island extraction in similar environments.

We already discussed the trends that we found pertaining to specifically adjunct or relative clause extraction above (viz. the type of adjunct clause typically involved in adjunct clause extraction, or the type of matrix verb featured in relative clause extraction). One feature shared by both types of islands is the type of extraction dependency found in naturally occurring extractions. Specifically, we found that all retrieved instances of relative and adjunct clause extraction involved topicalization or relativization from the island, whereas *wh*-extraction remained unattested. This observation reinforces the notion that island extraction possibilities appear to vary across different types of extraction dependency, with *wh*-extraction being more restricted than the other dependency forms (Kush et al. 2018, 2019; Bondevik et al. 2020; Sprouse et al. 2016; Abeillé et al. 2020). A difference between for instance relativization and *wh*-extraction could perhaps be derived from a pragmatic account of island constraints, as proposed by Chaves and Putnam (2020) (see also Goldberg 2006). As outlined above, Chaves and Putnam (2020) suggest that relative clauses and adjuncts are islands because they usually convey backgrounded information, and extraction from them thus draws attention to a referent that cannot be construed as sufficiently relevant for the main action, thereby violating a pragmatic relevance principle. However, if extraction is in the form of relativization, this relevance constraint can according to Chaves and Putnam (2020) be circumvented, since relatives "express assertions rather than backgrounded information" (p. 91). Similarly, Abeillé et al. (2020) provide an account according to which relativization from backgrounded constituents may be easier to accept than *wh*-extraction, since—unlike *wh*-extraction—relativization does not put the fronted element into focus and can thus avoid a conflict with the backgrounded status of the domain that is extracted from. This account could be extended to cover the apparent possibility of topicalization from some islands (in the MSc. languages). While topicalization can target focused elements, it is crucially also used frequently to front non-focal elements, and can thus pattern with relativization in not requiring the extracted element to be focused. One potential problem of this approach is that the adjunct clauses and relative clauses primarily featured in our attested examples are of a type that should already help circumvent the backgroundedness or relevance constraint. As we discussed in Section 3.3, all adjunct clauses featured in our sample express a cause or general condition for the (often psychological) condition in the matrix clause, and can thus be considered to express at-issue information for the main action (see Chaves and Putnam 2020, p. 91). Similarly, in Section 4.3 we observed that the relative clause extractions we found have in common that the matrix predicate is void of or very low in semantic content and thus allows for the relative clause to express new and relevant information, or the main point of utterance. Following a proposal along the lines of Chaves and Putnam (2020) or Abeillé et al. (2020), one would expect that it should not be problematic to extract an element from these kinds of adjunct or relative clauses even if the extraction dependency puts the fronted element into focus (as is the case with *wh*-questions), since the domain that it is extracted from is no longer backgrounded and thus should no longer constrain constructions that make the extracted element prominent in the discourse. This would leave unexplained the absence of *wh*-extraction in our sample.

A possible alternative explanation is that the relative and adjunct clauses in our examples, despite being less backgrounded, are still behaving like a type of *weak island* (Szabolcsi 2006), possibly because of a syntactic property. Weak islands are known to permit extraction in some configurations (e.g., if the extracted element is a discourse-linked argument), but not in others. Adjuncts and relative clauses in MSc. have previously been suggested to be weak islands based on their selective extraction behavior (e.g., Lindahl 2017; Müller 2019).

A prominent approach to weak islands is in terms of *featural Relativized Minimality* (Starke 2001; Rizzi 2013). According to featural RM, the acceptability of weak island extraction is dependent on the featural specification of the involved elements: An element can be extracted as long as the movement path does not cross an intervening element that is more richly specified than the extracted element, as this would lead to an intervention effect. Weak island effects caused by adjunct and relative clauses can be accounted for under featural RM by assuming that these clauses involve an operator in their left periphery that can cause said intervention effect (see e.g., Demirdache and Uribe-Etxebarria 2004; Bhatt and Pancheva 2006; Haegeman 2010, 2012) for evidence that conditional and temporal adjunct clauses are derived by operator movement). In order to derive differences between dependency types from a RM account, one would have to assume that topicalization and relativization differ from *wh*-extraction in that they both involve movement of an element that carries a richer featural specification than the operator present in relative or adjunct clauses. There are several proposals that group topicalization and relativization together, based on their close relation (e.g., Kuno 1976; Williams 2011; Abels 2012; Douglas 2016). Williams (2011, 2013) shows several ways in which relatives and topicalization pattern together against *wh*-questions, and Douglas (2016, p. 147) specifically suggests that topicalization and relativization share a feature (presumably for a discourse-linking property) that *wh*-questions do not have.

The additional feature that is assumed to be present in the types of phrases that can escape from weak islands is often taken to be a referential or *discourse-linking* (Pesetsky 1987) feature. This should also enable extraction by question formation from the island, as long as the extracted phrase is a lexically restricted *wh*-phrase of the type *which NP*, as these favor a D-linked interpretation. Nevertheless, extraction of complex *wh*-phrases is just as absent from our sample as extraction of bare *wh*-phrases. However, there have been suggestions that more fine-grained interpretive properties govern the possibility of weak island extraction: Starke (2001) and Baunaz (2011) observe that extraction from some weak islands seems to trigger presupposition of a specific referent that must be familiar to both speaker and hearer, rather than just a D-linked interpretation, where the extracted referent is expected to be part of a set of alternatives that is salient in the discourse. It is possible that topicalization and relativization fulfill these requirements for a specific reference more easily than question formation. In fact, Starke (2001) and Baunaz (2011) point out that questions which trigger the specific reading of the *wh*-phrase require a very rich context and are akin to echo-questions. This could then explain why *wh*-extraction from e.g., relative clauses is judged to be possible with constructed examples—especially with an echo-question reading (see Engdahl 1997; Lindahl 2017)—but occurs rarely in production, as the appropriate kinds of context can be constructed if necessary, but probably do not often occur in (written) natural language. We acknowledge the speculative nature of these remarks, but leave it to future research to investigate further in how far a more fine-grained approach to feature composition can account for potential asymmetries between different extraction dependencies.

As a final note, we would like to mention that we think these results have the potential to complement and inform future formal studies of island phenomena, in the sense that many acceptability studies aim to use stimuli that are modelled on naturally occurring data. However, our corpus study revealed for instance that most naturally attested island extractions in Danish and English involve relativization or (in Danish) topicalization, while *wh*-extraction remained unattested. In this regard, the naturally produced extractions

contrast with the test sentences used in most acceptability judgment experiments on island extractions, which—at least in English—have focused on testing extraction in the form of *wh*-movement. At the same time, further formal acceptability studies of island constructions e.g., in Danish, are also necessary to get a better understanding of the source of the apparent mismatches between experiments and naturally occurring production.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/languages7020125/s1.

Author Contributions: Conceptualization, C.M. and C.U.E.; methodology, C.M. and C.U.E.; formal analysis, C.M. and C.U.E.; investigation, C.M. and C.U.E.; data curation, C.M. and C.U.E.; writing—original draft preparation, C.M.; writing—review and editing, C.M. and C.U.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Independent Research Fund Denmark (grant DFF-9062-00047B).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the Supplementary materials.

Acknowledgments: We wish to thank Anne Mette Nyvad, Ken Ramshøj Christensen, and Sten Vikner for assistance with the design of the corpus studies and for providing valuable comments on this research. For further comments and suggestions, we would like to thank two anonymous reviewers as well as the editors for this *Languages* issue.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1] Superadditive effects have been taken to be a diagnostic for islandhood in the factorial design developed by Sprouse (2007), which factors out the influence of the presence of a long-distance dependency as well as of the presence of an island configuration on acceptability. A significant interaction between these two factors in the factorial design implies that they create a superadditive effect on acceptability, i.e. an effect that cannot be explained by the individual costs of having an extraction and an island structure.

[2] Relatedly, Kush et al. (2013) propose that the possibility of relative clause extraction in MSc. is tied to the matrix verb's ability to select a small clause; however, this proposal has been called into question by e.g., Christensen and Nyvad (2014), Müller (2015), and Heinat and Wiklund (2015).

References

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204: 104293. [CrossRef] [PubMed]

Abels, Klaus. 2012. The Italian left periphery: A view from locality. *Linguistic Inquiry* 43: 229–54. [CrossRef]

Allwood, Jens S. 1982. The complex NP constraint in Swedish. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabeth Engdahl and Eva Ejerhed. Stockholm: Almqvist & Wiksell, pp. 15–32.

Anward, Jan. 1982. Basic Swedish. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabeth Engdahl and Eva Ejerhed. Stockholm: Almqvist & Wiksell, pp. 47–75.

Baunaz, Lena. 2011. *The Grammar of French Quantification*. Dordrecht and New York: Springer.

Bhatt, Rajesh, and Roumyana Pancheva. 2006. Conditionals. In *The Blackwell Companion to Syntax*. Edited by Martin Everaert and Henk van Riemsdijk. Oxford and Boston: Blackwell Publishing, pp. 638–87.

Boeckx, Cedric. 2012. *Syntactic Islands*. Cambridge: Cambridge University Press.

Bondevik, Ingrid. 2018. Investigating the universality of adjunct islands through formal acceptability experiments. A comparative study of English and Norwegian. Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway.

Bondevik, Ingrid, Dave Kush, and Terje Lohndal. 2020. Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics* 44: 223–54. [CrossRef]

British National Corpus. 2001. The Version 2 (BNC World). Distributed by Oxford University Computing Services on Behalf of the BNC Consortium. Available online: http://www.natcorp.ox.ac.uk/ (accessed on 19 November 2021).

Casalicchio, Jan. 2016. Pseudo-relatives and their left-periphery. A unified account. In *Romance Languages and Linguistic Theory 10: Selected Papers from 'Going Romance' 28, Lisbon*. Edited by Ernestina Carrilho, Alexandra Fiéis, Maria Lobo and Sandra Pereira. Philadelphia: Benjamins, pp. 23–42.
Cattell, Ray. 1976. Constraints on movement rules. *Language* 52: 18–50. [CrossRef]
Chaves, Rui P., and Michael T. Putnam. 2020. *Unbounded Dependency Constructions*. Oxford: Oxford University Press.
Chomsky, Noam. 1986. *Barriers*. Cambridge: MIT Press.
Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2014. On the nature of escapable relative islands. *Nordic Journal of Linguistics* 37: 29–45. [CrossRef]
Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2022. The Island Is Still There: Experimental Evidence for the Inescapability of Relative Clauses in English. *Studia Linguistica* 2022: 1–25. [CrossRef]
Chung, Sandra, and James McCloskey. 1983. On the interpretation of certain island facts in GPSG. *Linguistic Inquiry* 14: 704–13.
Corpus of Contemporary American English. 2008. Available online: http://corpus.byu.edu/coca/ (accessed on 19 November 2021).
Culicover, Peter W. 1999. *Syntactic Nuts: Hard Cases, Syntactic Theory and Language Acquisition*. Oxford: Oxford University Press.
Davies, Mark. 2013. Corpus of Global Web-Based English: 1.9 Billion Words from Speakers in 20 Countries (GloWbE). Available online: https://corpus.byu.edu/glowbe/ (accessed on 19 November 2021).
Deane, Paul. 1991. Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics* 2: 1–63. [CrossRef]
Demirdache, Hamida, and Myriam Uribe-Etxebarria. 2004. The syntax of time adverbs. In *The Syntax of Time*. Edited by Jaqueline Guéron and Jacqueline Lecarme. Cambridge: MIT Press, pp. 143–80.
Douglas, Jamie A. 2016. The Syntactic Structures of Relativisation. Ph.D. dissertation, University of Cambridge, Cambridge, UK.
Engdahl, Elisabeth. 1997. Relative clause extractions in context. *Working Papers in Scandinavian Syntax* 60: 59–86.
Ernst, Thomas. 2022. The adjunct condition and the nature of adjuncts. *The Linguistic Review* 39: 85–128. [CrossRef]
Erteschik-Shir, Nomi. 1973. On the Nature of Island Constraints. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
Erteschik-Shir, Nomi. 1982. Extrability in Danish and the pragmatic principle of dominance. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabeth Engdahl and Eva Ejerhed. Stockholm: Almqvist & Wiksell, pp. 175–91.
Goldberg, Adele. 2006. *Constructions at Work. The Nature of Generalization in Language*. New York: Oxford University Press.
Goldberg, Adele. 2013. Backgrounded constituents cannot be extracted. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 221–38.
Haegeman, Liliane. 2004. The syntax of adverbial clauses and its consequences for topicalisation. In *Current Studies in Comparative Romance Linguistics*. Edited by Martine Coene, Gretel De Cuyper and Yves D'hulst. Antwerp: University of Antwerp, pp. 61–90.
Haegeman, Liliane. 2010. The internal syntax of adverbial clauses. *Lingua* 120: 628–48. [CrossRef]
Haegeman, Liliane. 2012. *Adverbial Clauses, Main Clause Phenomena, and the Composition of the Left Periphery*. Oxford: Oxford University Press.
Hagström, Björn. 1976. Om satsflätans förutsättningar. En skiss [On the conditions for subextraction. A sketch]. In *Nordiska studier i filologi och lingvistik. Festskrift tillägnad Gösta Holm på 60-årsdagen den 8 juli 1976*. Edited by Lars Svensson, Anne Marie Wieselgren and Åke Hansson. Lund: Studentlitteratur, pp. 138–50.
Hansen, Erik, and Lars Heltoft. 2011. *Grammatik over det danske sprog [Danish Reference Grammar]*. København: Det danske Sprog-og Litteraturselskab, vol. 3.
Heinat, Fredrik, and Anna-Lena Wiklund. 2015. Scandinavian relative clause extraction. Apparent restrictions. *Working Papers in Scandinavian Syntax* 94: 36–50.
Hofmeister, Philip, and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415. [CrossRef]
Huang, James C.-T. 1982. Logical Relations in Chinese and the Theory of Grammar. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
Jensen, Anne. 1998. Knudekonstruktioner—En syntaktisk, semantisk og pragmatisk analyse av sætningsknuder i dansk. Master's thesis, Københavns Universitet, Copenhagen, Denmark.
Jensen, Anne. 2002. Sætningsknuder i dansk [Sentence intertwining in Danish]. *NyS—Nydanske Studier & Almen Kommunikationsteori* 29: 105–24.
Johnston, Michael. 1994. The Syntax and Semantics of Adverbial Adjuncts. Ph.D. dissertation, University of California, Santa Cruz, CA, USA.
KorpusDK. 2021. Available online: https://ordnet.dk/korpusdk/ (accessed on 19 November 2021).
Kuno, Susumu. 1976. Subject, theme, and the speaker's empathy: A re-examination of relativization phenomena. In *Subject and Topic*. Edited by Charles N. Li. New York: Academic Press, pp. 417–44.
Kush, Dave, Akira Omaki, and Norbert Hornstein. 2013. Microvariation in Islands? In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 239–65.
Kush, Dave, Charlotte Sant, and Sunniva Briså Strætkvern. 2021. Learning Island-insensitivity from the input: A corpus analysis of child- and youth-directed text in Norwegian. *Glossa* 6: 105. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian Wh-extraction. *Natural Language & Linguistic Theory* 36: 743–79. [CrossRef]

Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95: 393–420. [CrossRef]
Lindahl, Filippa. 2010. Spetsställda led och rematiska relativer. En korpusstudie av satsfläta med presenteringsomskrivning/existentialsats. Master's thesis, University of Gothenburg, Gothenburg, Sweden.
Lindahl, Filippa. 2017. Extraction from Relative Clauses in Swedish. Ph.D. dissertation, University of Gothenburg, Gothenburg, Sweden.
Löwenadler, John. 2015. Relative clause extraction: Pragmatic dominance, processing complexity and the nature of crosslinguistic variation. *Nordic Journal of Linguistics* 38: 37–65. [CrossRef]
MacWhinney, Brian, and Johannes Wagner. 2010. Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprächsforschung* 11: 154–73.
Maling, Joan, and Annie Zaenen. 1982. A phrase structure account of Scandinavian extraction phenomena. In *The Nature of Syntactic Representation*. Edited by Pauline Jacobson and Geoffrey K. Pullum. Dordrecht: Reidel, pp. 229–82.
Manzini, Maria R. 1992. *Locality: A Theory and Some of Its Empirical Consequences*. Cambridge: MIT Press.
McCawley, James D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53: 99–149. [CrossRef]
Michel, Dan, and Grant Goodall. 2013. Finiteness and the nature of island constraints. In *Proceedings of GLOW in Asia IX 2012: The Main Session*. Edited by Nobu Goto, Koichi Otaki, Atsushi Sato and Kensuke Takita. Tsu: Mie University, Japan, pp. 187–97.
Müller, Christiane. 2015. Against the Small Clause Hypothesis: Evidence from Swedish relative clause extractions. *Nordic Journal of Linguistics* 38: 67–92. [CrossRef]
Müller, Christiane. 2017. Extraction from Adjunct Islands in Swedish. *Norsk Lingvistisk Tidsskrift* 35: 67–85.
Müller, Christiane. 2019. Permeable Islands: A Contrastive Study of Swedish and English Adjunct Clause Extraction. Ph.D. dissertation, University of Lund, Lund, Sweden.
Nyvad, Anne Mette, Christiane Müller, and Ken Ramshøj Christensen. 2022. Too good to be true: The nonuniformity of adjunct clause extraction in English. *Languages*, *under review*.
Nyvad, Anne Mette, Ken Ramshøj Christensen, and Sten Vikner. 2017. CP-recursion in Danish: A cP/CP-analysis. *The Linguistic Review* 34: 449–77. [CrossRef]
Pesetsky, David. 1987. *Wh*-in-situ: Movement and unselective binding. In *The Representation of (In)Definiteness*. Edited by Eric J. Reuland and Alice ter Meulen. Cambridge: MIT Press, pp. 98–129.
Poole, Ethan. 2017. Movement and the Semantic Type of Traces. Ph.D. dissertation, University of Massachusetts, Amherst, MA, USA.
Poulsen, Mads. 2008. Acceptability and processing of long-distance dependencies in Danish. *Nordic Journal of Linguistics* 31: 73–107. [CrossRef]
Rizzi, Luigi. 2013. Locality. *Lingua* 130: 169–86. [CrossRef]
Ross, John R. 1967. Constraints on Variables in Syntax. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
Sawada, Miyuki, and Richard K. Larson. 2004. Presupposition & root transforms in adjunct clauses. In *Proceedings of NELS 34*. Edited by Matthew Wolf and Keir Moulton. Amherst: GLSA, pp. 517–28.
Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117: 1034–56. [CrossRef]
Sprouse, Jon. 2007. A Program for Experimental Syntax: Finding the Relationship between Acceptability and Grammatical Knowledge. Ph.D. dissertation, University of Maryland, College Park, MD, USA.
Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34: 307–44. [CrossRef]
Sprouse, Jon, Matthew W. Wagers, and Colin Phillips. 2012. A test of the relation between working memory capacity and syntactic island effects. *Language* 88: 82–123. [CrossRef]
Starke, Michal. 2001. Move Dissolves into Merge: A Theory of Locality. Ph.D. dissertation, University of Geneva, Geneva, Switzerland.
Szabolcsi, Anna. 2006. Strong vs. weak islands. In *The Blackwell Companion to Syntax*. Edited by Martin Everaert and Henk van Riemsdijk. Oxford: Blackwell, pp. 479–531.
Taraldsen, Knut T. 1982. Extraction from relative clauses in Norwegian. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabeth Engdahl and Eva Ejerhed. Stockholm: Almqvist & Wiksell, pp. 205–22.
Teleman, Ulf, Staffan Hellberg, and Erik Andersson. 1999. *Svenska Akademiens Grammatik [Swedish Reference Grammar]*. Stockholm: Svenska akademien, vol. 4.
Truswell, Robert. 2007. Locality of Wh-Movement and the Individuation of Events. Ph.D. dissertation, UCL, London, UK.
Truswell, Robert. 2011. *Events, Phrases and Questions*. Oxford: Oxford University Press. [CrossRef]
Tutunjian, Damon, Fredrik Heinat, Eva Klingvall, and Anna-Lena Wiklund. 2017. Processing relative clause extractions in Swedish. *Frontiers in Psychology* 8: 2118. [CrossRef]
Van Valin, Robert D. 1994. Extraction restrictions, competing theories and the argument from the poverty of the stimulus. In *The Reality of Linguistic Rules*. Edited by Susan D. Lima, Roberta L. Corrigan and Gregory K. Iverson. Philadelphia: Benjamins, pp. 243–59.
Van Valin, Robert D. 1996. Toward a functionalist account of so-called extraction constraints. In *Complex Structures. A Functionalist Perspective*. Edited by Betty Devrient, Louis Goossens and Johan van der Auwera. Berlin and New York: De Gruyter, pp. 29–60.
Van Valin, Robert D. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Vincent, Jake Wayne. 2021. Extraction from Relative Clauses: An Experimental Investigation into Variable Island Effects in English. Ph.D. dissertation, University of California, Santa Cruz, CA, USA.
Wiklund, Anna-Lena, Fredrik Heinat, Eva Klingvall, and Damon Tutunjian. 2017. An acceptability study of long-distance extractions in Swedish. In *Language Processing and Disorders*. Edited by Linda Escobar, Vincenç Torrens and Teresa Parodi. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 103–20.
Williams, Edwin S. 2011. *Regimes of Derivation in Syntax and Morphology*. London: Routledge.
Williams, Edwin S. 2013. Generative Semantics, Generative Morphosyntax. *Syntax* 61: 77–108. [CrossRef]

languages

MDPI

Article

Too True to Be Good? The Non-Uniformity of Extraction from Adjunct Clauses in English

Anne Mette Nyvad *, Christiane Müller and Ken Ramshøj Christensen

Department of English, School of Communication and Culture, Aarhus University, 8000 Aarhus, Denmark
* Correspondence: amn@cc.au.dk

Abstract: Adjunct clauses are traditionally assumed to be strong islands for extraction across languages. However, the universal island status of adjunct clauses has been challenged by studies showing that extraction is possible from finite adjunct clauses in the Mainland Scandinavian languages. The possibility of extraction in these languages appears to be affected by various factors, including the type of adjunct clause, the type of extraction dependency, and the presence of contextual facilitation. These findings call for a re-evaluation of the islandhood of adjunct clauses in English. We conducted an acceptability judgment study on relativization from three types of finite adjunct clauses in English (*if*-, *when*-, and *because*-clauses) in the presence of supporting context. We found that the three clause types showed rather non-uniform acceptability patterns: extraction from *when*- and *because*-clauses both yielded significantly lower ratings than extraction from *if*-clauses, which patterned with non-island *that*-clauses. Our results suggest that at least for relativization, *if*- and *when*-adjuncts are not invariably strong islands in English, and that extra-grammatical factors may be key in understanding island structures traditionally assumed to be purely syntactic in nature.

Keywords: syntax; islands; acceptability; grammaticality; satiation; variation

Citation: Nyvad, Anne Mette, Christiane Müller, and Ken Ramshøj Christensen. 2022. Too True to Be Good? The Non-Uniformity of Extraction from Adjunct Clauses in English. *Languages* 7: 244. https://doi.org/10.3390/languages7040244

Academic Editors: Juana M. Liceras and Raquel Fernández Fuertes

Received: 22 April 2022
Accepted: 5 September 2022
Published: 21 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

One of the defining characteristics of human language is that an element in a sentence can appear in a different position than the one in which it is interpreted, e.g., in long-distance extraction:

1. Who do you think [that I saw __]?

In this example, the question word *who* has been displaced to the front of the matrix clause, but it is interpreted as the direct object of the main verb in the embedded clause. The link between these two positions is called a filler-gap dependency and such dependencies are in principle unbounded. However, Ross (1967) identified a range of structures that (to varying degrees) block extraction which he termed "islands". It has since become common to distinguish between weak (or selective) and strong (or absolute) islands (e.g., Kluender 1998; Szabolcsi 2006; Szabolcsi and Lohndal 2017), depending on the degree to which it is possible to extract from such structures.

Island effects are widespread across the languages of the world and they have been used as an argument for the existence of innate linguistic principles. The language-acquiring child is allegedly not exposed to these types of extractions and the lack of negative evidence has been taken to suggest that certain constraints on filler-gap dependencies are part of universal grammar, either as principles or parameters.

Effects of repeated exposure (satiation, priming, and/or training effects) have been widely attested in the literature on island structures (Chaves and Putnam 2020; Christensen et al. 2013; Snyder 2000, 2022). Such effects may suggest that the constructions in question are in fact not illicit, but rather that they are unusual and/or highly complex, giving rise to initial processing difficulty and resulting in decreased acceptability, which is then ultimately ameliorated as speakers adapt to them.

In the syntax literature, finite adjunct clauses have traditionally been assumed to be strong islands cross-linguistically, as exemplified in (2) from Szabolcsi (2006, p. 481):

2. *Which topic did you leave [because Mary talked about __]?

The ungrammaticality of (2) has been accounted for under the *Condition on Extraction Domain* (CED, Huang 1982, p. 505). This constraint states that a phrase can only be extracted out of a domain that is properly governed, and hence both subjects and adjuncts should be island environments (cf. the Empty Category Principle, Chomsky 1981, p. 205ff; Haegeman 1994, p. 442). The CED has been thought to be an innate, universal principle (along with a number of other island constraints, subsequently subsumed under e.g., subjacency, Chomsky 1973), given the learnability problem imposed by the purported lack of negative evidence in the input during language acquisition. However, counterexamples abound in the Mainland Scandinavian (MSc.) languages. Recent studies on Norwegian (Kush et al. 2019; Bondevik et al. 2020) have found that topicalization (but not *wh*-movement) out of conditional adjunct clauses headed by *om* ('if') appears to be possible. In addition, Müller (2017) found that topicalization from adjunct clauses headed by *om* ('if') and *efter* ('after') in Swedish yielded acceptability ratings on the upper end of the scale, provided that the matrix and the adjunct clause event can be interpreted to be in a causal, coherent relation e.g., (3a), rather than in a purely temporal one e.g., (3b).

3. a. Det där röda vinet mådde jag lite illa
this there red wine.the felt I a.little sick
[efter att jeg hade druckit ___ sist].
after that I had drunk last
'I felt a little sick after I had drunk that red wine last time.'

b. Det röda kan vi ju gå ut på stan
the red can we PRT[1] *go out on town.the*
[efter att vi har druckit ___].
after that we have drunk
'We can go out on the town after we have drunk the red wine.'

However, other types of adjunct clauses—specifically, clauses introduced by *fordi* 'because' in Norwegian and *eftersom* 'because' in Swedish—were found to be much less permissive in terms of extraction in these studies in the sense that most participants rated topicalization from them on the low end of the scale (Bondevik et al. 2020; Müller 2017). This has given rise to conjectures that different types of adjunct clauses, or adjunct clauses headed by different complementizers, may vary with regard to extraction possibilities (Bondevik 2018, p. 201; Bondevik et al. 2020; Müller 2017, 2019). Moreover, the studies by Kush et al. (2018, 2019) and Bondevik et al. (2020), among others, suggest that topicalization from adjunct clauses is easier to accept than question formation by *wh*-extraction, indicating that the acceptability of extraction might also depend on the type of extraction (see also Lindahl 2017; Müller 2019; and Abeillé et al. 2020 for suggestions along these lines).

Another factor that has been suggested to be important for the felicity of at least some island extractions in MSc. in formal acceptability studies is whether the stimulus sentences are preceded by one or more sentences providing a facilitating context. More specifically, some of the formal studies of adjunct or relative clause extraction in MSc. have yielded acceptability ratings that are unexpectedly low, given the examples provided in the literature and the informal observations reporting such structures to be acceptable in MSc. (e.g., Christensen and Nyvad 2014; Kush et al. 2018). It has been suggested that these unexpectedly low ratings can be partly explained by the lack of contextual cues in formal settings (e.g., Tutunjian et al. 2017; Wiklund et al. 2017, p. 207; Kush et al. 2018; Müller 2019, p. 209; Bondevik et al. 2020). Contextual cues may be required for the felicity of at least some of these extractions if for instance a context is necessary to license certain interpretational properties of the extracted phrase, such as D(iscourse) linking or specificity (Pesetsky 1987; Szabolcsi 2006). The licensing of such interpretational properties has been claimed to be relevant for the felicity of weak island extractions (see e.g., Starke 2001; Szabolcsi 2006;

Abrusán 2014). In a recent study on Norwegian, Kush et al. (2019) found support for the beneficial effect of context on ratings for at least some adjunct clause extractions: A significant island effect (in terms of the factorial definition of island effects developed by Sprouse 2007) was observed for decontextualized extraction from conditional adjuncts, but no significant island effect occurred once a contextual preamble was added to the stimuli sentences. In fact, the participants in this study rejected sentences involving contrastive topicalization from complement clauses, uncontroversially considered to be grammatical, presumably at least in part because they were presented in vacuo. More general evidence in favor of the facilitative role of supporting context for the comprehension of fronted elements is presented by Engelkamp and Zimmer (1983), Kristensen et al. (2014), and Lau et al. (2020), but see Bernardy et al. (2018). The question remains whether English sentences such as (2) can be ameliorated by being presented in a facilitating context.

Kush et al. (2018) found significant inter-individual variation relating to the acceptability judgements of sentences involving *wh*-movement out of embedded polar questions in Norwegian (headed by *om* 'if'), including a large number of participants not exhibiting any island sensitivity to this type of extraction. Kush et al. (2018) conclude that this pattern of inter-participant variation is incompatible with embedded questions being syntactic islands in Norwegian and they argue that the variability in acceptability is actually indicative of extra-syntactic factors being at play. Similarly, Bondevik et al. (2020) and Kush et al. (2019) found both inter- and intra-participant variation as well as variation between and within adjunct types in Norwegian.

The above-mentioned studies (Bondevik et al. 2020; Kush et al. 2018, 2019; Müller 2017) indicate that adjunct clauses in the MSc. languages display a non-uniform behavior when it comes to their island sensitivity, and that the possibility of extraction is affected by various factors, including the type of adjunct clause, the type of extraction dependency, contextual facilitation, and semantic coherence[2]. These findings seem to call for a re-evaluation of the situation in English regarding adjunct islandhood. To date, only few formal acceptability studies of extraction from finite adjunct clauses in English exist, and almost all of them focus on *wh*-extraction from only one type of adjunct clause and do not provide any contextual facilitation (e.g., Sprouse 2009; Sprouse et al. 2012; Michel and Goodall 2013). Though Sprouse et al. (2016) also investigated extraction by relativization from *if*-adjuncts, they did not compare it to extraction from other types of adjunct clauses. In order to adequately evaluate whether English finite adjunct clauses really are uniformly strong islands, as reported in the traditional literature (and to allow for a fair comparison between English and MSc. adjunct islands), an investigation that tests adjunct clause extraction in English controlling for the above-mentioned factors (type of adjunct, dependency type, presence/absence of supporting context, and coherence) is necessary. In this paper, we present a study that aims to fill (parts of) this gap, by testing extraction in the form of relativization from three different types of finite adjunct clauses in English (*if*-, *when*-, and *because*-clauses) in the presence of supporting context.

We chose to test extraction in the form of relativization rather than topicalization since topicalization is a fairly marked structure in English compared to the MSc. languages (e.g., Engdahl 1997; Poole 2017, p. 15). There are some indications that relativization behaves on a par with topicalization in that both seem to facilitate extraction from certain islands, compared to *wh*-movement. Sprouse et al. (2016) compared *wh*-extraction and relativization out of adjunct clauses in English and found that relativization from *if*-clauses does not result in island effects in terms of the factorial definition of islands (although ratings for both relativization and question formation remained relatively low). Recently, Abeillé et al. (2020) showed that relativization (but not question formation) from subject islands in English as well as in French received ratings on par with grammatical controls. Moreover, there are indications that spontaneously produced cases of adjunct clause extractions primarily feature relativization from the adjunct: to see whether any authentic examples of adjunct island extraction can be found in English, Müller and Eggers (2022) conducted an exploratory corpus study on adjunct extraction using the *Corpus of Contemporary American English* (COCA; Davies

2008). All cases of adjunct clause extraction found in naturalistic English in their study involved relativization from the island, such as the examples in (4), both from COCA.

4. a. Many of the exercises are ones that I would be surprised if even 1 percent of healthy women can do.
   b. Now, those are things that I feel very warm when I look at, and I wouldn't want to live in a house that they—a house that didn't have room for those.

On a purely syntactic account, the level of acceptability of extraction out of finite adjunct clauses should be uniform across adjunct types, given that they have the same syntactic structure and that they are assumed to be adjoined to the matrix clause in the same way. Furthermore, gradient acceptability of extraction from adjunct clauses is not predicted by the CED either.

As an alternative to purely syntactic accounts of island constraints, Chaves and Putnam (2020) propose that a host of islands traditionally assumed to be of syntactic nature are instead "Relevance Islands": the filler must be "relevant for the main action that the proposition conveys" and if it is not, unacceptability ensues as a natural consequence, without the need for postulating constraints in the syntax (Chaves and Putnam 2020, p. 120; in line with other semantic or discourse-pragmatic accounts such as Erteschik-Shir 1973; Deane 1991). Prima facie, even a semantico-pragmatic account of island effects such as the one proposed by Chaves and Putnam would predict uniformly low acceptability of adjunct clause extraction since adjunct clauses are typically used to express backgrounded or presupposed information (see also e.g., Van Valin 1994; Erteschik-Shir 2006; Goldberg 2013, p. 203 for accounts in a similar vein).

2. Predictions

In light of the counterexamples to supposed universal constraints found in the MSc. languages, Christensen and Nyvad (2019, 2022) re-examined *wh*-islands and relative clause extraction in English in order to investigate whether the same island-insensitivity and graded acceptability could also be demonstrated for English. However, the results support the hypothesis that the syntactic configurations in question (embedded *wh*-questions and relative clauses) are strong islands in this language, as is standardly assumed in the syntactic literature (but see Vincent et al. 2022). Given this compatibility of experimental findings in English with the standard assumptions in the syntax literature (the *Wh*-Island Constraint and the Complex NP Constraint, respectively), our predictions regarding extraction from adjunct clauses are based on the CED, which treats all adjunct clauses as a coherent class in terms of islandhood, potentially ruling out any scope for variation:

Prediction 1. *Adjunct clauses are strong islands, i.e., their acceptability level is consistently low.*

Prediction 2. *There is no variation in acceptability as a function of adjunct clause type.*

Prediction 3. *There is no between-participant variation.*

Prediction 4. *There is no trial effect. The acceptability of extraction is not positively correlated with repeated exposure over time in the experiment, regardless of the type of adjunct clause.*

We included subject islands and coordinate structure islands as fillers and as points of comparison, given that these structures are also assumed to be strong islands in English, but with different strengths. The Subject Condition (Ross 1967; Chomsky 1973) has been proposed not to be active across all types of constructions: Kush et al. (2018, 2019) found substantial island effects in the acceptability rating of constructions involving movement out of subject phrases. However, extraction from subjects has been demonstrated to be grammatical in certain instances, depending on referential processing (Culicover and Winkler 2022; Culicover et al. 2022), discourse function (Abeillé et al. 2020), and repeated exposure (Chaves and Putnam 2020, p. 221). The Coordinate Structure Constraint, on the other hand, appears to hold cross-linguistically, and no robust counterexamples seem to

have been identified (Chaves and Putnam 2020, p. 72).[3] This way, we had two different points of comparison in terms of patterns of ungrammaticality.

3. Materials and Methods

Our design was based on a 2 × 2 stimulus design with the structure outlined in Table 1, with ±Island and ±Extraction as the two factors; this is similar to the factorial design introduced by Sprouse (2007) and subsequently employed by e.g., Sprouse et al. (2011, 2016) and Bondevik et al. (2020). In addition, our design had a Complementizer factor with four different levels: *that*, *if*, *when*, and *because*. *That*-clauses are not assumed to be islands, since they are complement clauses, whereas the other three types introduce adjunct clauses. This design gave us eight target types (see Table 1), which we used as the levels for a Type factor for our statistical model in order to test for interactions between extraction and the four clause types.

Table 1. Experimental design.

	Complementizer	**[−Extraction]**	**[+Extraction]**
Non-island	*That*	Type 1	Type 5
Islands	*If*	Type 2	Type 6
	When	Type 3	Type 7
	Because	Type 4	Type 8

Each target stimulus sentence was preceded by a short facilitating context. We constructed the materials in such a way that all eight types could be preceded by the same context, as shown in (5):

5. **Context**: In the latest workout routine I designed for Emma, I really wanted to make it impossible for her and included another set of particularly brutal pull-ups.
   Non-island structure, [-Extraction]:
   a. It's obvious that I was surprised **that** she actually completed this exercise.
   Island structure, [-Extraction]:
   b. It's obvious that I would be surprised **if** she actually completed this exercise.
   c. It's obvious that I was surprised **when** she actually completed this exercise.
   d. It's obvious that I was surprised **because** she actually completed this exercise.
   Non-island structure, [+Extraction]:
   e. This is the exercise that I was surprised **that** she actually completed __.
   Island structure, [+Extraction]:
   f. This is the exercise that I would be surprised **if** she actually completed __
   g. This is the exercise that I was surprised **when** she actually completed __.
   h. This is the exercise that I was surprised **because** she actually completed __.

Distinct from the factorial design introduced by Sprouse (2007) and subsequently employed by e.g., Sprouse et al. (2011, p. 201; 2016), Bondevik et al. (2020), the main matrix clauses of the stimulus sentences are not minimally different, but they are as minimally different as they can be. The sentences without extraction (the baseline versions) all begin with *It's obvious that* . . . or *It's clear that* . . . , whereas the sentences with extraction all begin with *This is the X that* . . . (where *the X* is the extracted filler). The stimulus set was created this way because we prioritized having relativization out of the adjunct clauses (see Section 1 above). There is no direct unextracted counterpart to relativization that our [+Extraction] sentences could be compared to, but the *It's obvious that* . . . or *It's clear that* . . . constructions used for the matrix in our baseline sentences are as similar as possible in terms of length and complexity to the cleft structure used in the relativization condition. Note that both matrix structures add very little to the overall structure and meaning (we used pronouns and copula *be*). We acknowledge that this makes the ±Extraction contrast not minimally different, but there are no a priori reasons to suspect that these small differences would affect the acceptability ratings.

As illustrated in (5) and in Table 1, we wanted to compare three different types of adjunct islands: *if*-clauses, *when*-clauses and *because*-clauses. The adjunct clauses and non-island *that*-clauses were embedded under adjectival psych-predicates such as *be surprised* or *be happy*, which can be expected to trigger a causal and thus semantically coherent reading of the matrix and embedded event. The presence of a causal, coherent relation between the matrix and the adjunct clause has previously been shown to facilitate extraction from adjuncts (Truswell 2011; Tanaka 2015; Müller 2019). In order to allow for a felicitous use of the *if*-clauses, the verbal structure in the matrix clause was slightly altered in the conditions involving *if*-adjuncts (5b,f) compared to the other conditions: instead of past tense *was*, the structure *would be* + psych-adjective was used, signaling the hypothetical.

Twenty-four sets of items of the type in (5) were constructed and distributed across eight lists in a Latin square design. In this way, each list contained only one version of the same scenario; since each participant saw only one list, each participant was exposed to three tokens of each type. The order of sentences on each list was randomized. As mentioned in the introduction, we also added two types of fillers: eight sets involving extraction from NP subjects, filler 1, as in (6), and eight sets involving extraction from coordinate structures, filler 2, as in (7). The NP subjects and the coordinate structures were embedded in the same four clause types that we used in target items, i.e., non-island *that*-clauses, and adjunct clauses introduced by *if*, *when*, and *because* (see Table 1). Each of these conditions with extraction was compared to the corresponding version without extraction, resulting in eight control conditions that could easily be compared to our target items. (The entire stimulus set is available online, see Data Availability Statement).

6. **Filler 1.**
Context: My team has developed a COVID-19 vaccine in record time, and I think we deserve some recognition.
Subject island + non-island that-clause, [-Extraction]:
a. It's clear that we were pleased **that** our vaccine against this virus finally got the Nobel Prize.
Subject island + adjunct island, [-Extraction]:
b. It's clear that we would be pleased **if** our vaccine against this virus finally got the Nobel Prize.
c. It's clear that we were pleased **when** our vaccine against this virus finally got the Nobel Prize.
d. It's clear that we were pleased **because** our vaccine against this virus finally got the Nobel Prize.
Subject island + non-island that-clause, [+Extraction]:
e. This is the virus that we were pleased **that** our vaccine against __ finally got the Nobel Prize.
Subject island + adjunct island, [+Extraction]:
f. This is the virus that we would be pleased **if** our vaccine against __ finally got the Nobel Prize.
g. This is the virus that we were pleased **when** our vaccine against __ finally got the Nobel Prize.
h. This is the virus that we were pleased **because** our vaccine against __ finally got the Nobel Prize.

. **Filler 2.**
Context: I had promised a friend to watch his pets for a week. Unfortunately, I accidentally left the front door open just a bit too long and both his cat and his prize-winning show dog ran out.
Coordinate structure island + non-island that-clause, [-Extraction]:
a. It's obvious that I was ashamed **that** I actually lost both the cat and this dog in one day.
Coordinate structure island + adjunct island, [-Extraction]:
b. It's obvious that I would be ashamed **if** I actually lost both the cat and this dog in one day.
c. It's obvious that I was ashamed **when** I actually lost both the cat and this dog in one day.
d. It's obvious that I was ashamed **because** I actually lost both the cat and this dog in one day.
Coordinate structure island + non-island that-clause, [+Extraction]:
e. This is the dog that I was ashamed **that** I actually lost both the cat and __ in one day.
Coordinate structure island + adjunct island, [+Extraction]:
f. This is the dog that I would be ashamed **if** I actually lost both the cat and __ in one day.
g. This is the dog that I was ashamed **when** I actually lost both the cat and __ in one day.
h. This is the dog that I was ashamed **because** I actually lost both the cat and __ in one day.

Participants were instructed to base their rating only on the sentence following the context. The context in our stimuli was constructed so that it triggered a cohesive interpretation of the extracted DP in the discourse in the [+Extraction] conditions.

Participants were recruited via professional connections and through various social media platforms and were randomly assigned to one of the eight presentation lists. They rated the sentences using an online questionnaire created with Google Forms. Judgments were given on a seven-point Likert scale ranging from 1 = 'completely unacceptable' to 7 = 'completely acceptable'. The participants had the opportunity to leave comments on every test sentence.

4. Results

A total of 235 native speakers of English (188 female, 41 male, 6 other) participated in the experiment on a voluntary basis. Participant age ranged from 18 to 74 years (mean = 33.19, SD = 10.79). The level of education in years (including e.g., primary and secondary school, college, university, PhD, etc.) ranged from 9 to 30 years (mean = 18.00, SD = 3.19). The number of participants per list ranged between 24 and 39 (list 1 = 25, list 2 = 28, list 3 = 30, list 4 = 29, list 5 = 39, list 6 = 32, list 7 = 24, list 8 = 28).

Prior to analysis, we removed one of the target sets (a set of 8 sentences with the same context) because it turned out the *that* [−Extraction] sentence (as in (5a) above) was pragmatically illicit. Since we wanted to use *that* [−Extraction] as baseline for the island effect size measure (i.e. the DD-score in Figure 4 below), including this in the analysis would skew the result. Table 2 contains the mean acceptability scores for all the conditions.

Table 2. Mean acceptability scores on a 7-point scale. [±Ex] = ±Extraction.

		Target Sentences		Filler 1		Filler 2	
		Acceptability	**SD**	**Acceptability**	**SD**	**Acceptability**	**SD**
That	[−Ex]	5.90	1.41	5.81	1.45	5.58	1.42
If	[−Ex]	6.27	1.11	5.83	1.42	5.77	1.27
When	[−Ex]	6.14	1.24	5.99	1.35	5.88	1.19
Because	[−Ex]	5.82	1.39	5.77	1.51	5.51	1.55
That	[+Ex]	4.97	1.72	2.78	1.69	1.91	1.23
If	[+Ex]	4.97	1.76	2.80	1.63	1.83	1.29
When	[+Ex]	3.92	1.84	2.37	1.55	1.86	1.32
Because	[+Ex]	3.39	1.74	2.11	1.42	1.70	1.19

The data was analyzed using *R* version 4.1.2 (R Core Team 2021) with the *lmerTest Package* (Kuznetsova et al. 2017) for mixed effects models and the *MASS Package* (Venables and Ripley 2002) for sliding, pair-wise contrasts of the type factor.

Extraction from adjunct clauses (target sentences as in (5) above), extraction from embedded subjects (filler type 1, as in (6)), and extraction from the second conjunct in a coordinate structure (filler 2, as in (7)) were analyzed using separate mixed-effects models. (We had no prior hypotheses about statistical differences between the three, and since the three sentence types were not minimally different, direct comparisons would also be inappropriate).

All plots were made in R using *ggplot2* (Wickham 2016) and *gridExtra* (Augie 2016).

4.1. Target Type: Adjunct Islands

The model contained acceptability as an outcome variable and type as a predictor (with sliding contrasts) and random intercepts for participant and item, and random slopes for type and trial by participant, and random slopes for trial by item. The summary of the results is provided in Table 3.

The results and sliding pairwise comparisons for the eight target sentence types are also shown in Figure 1A. The acceptability ratings for the four baseline conditions [−Ex] all had a rating above 5 on the 7-point scale. There was a significant drop in acceptability between the [−Ex] and [+Ex] types, indicating a significant negative main effect of extraction, [−Ex] > [+Ex], as shown by the contrast between *because* [−Ex] and *that* [+Ex] ($p < 0.001$, see also Figure 1). (Running the same model without sliding contrasts showed significant differences between *that* [−Ex] and *that* [+Ex] ($p < 0.001$) as well as between *that* [−Ex] and *if/when* [−Ex], while the difference between *that* [−Ex] and *because*

[−Ex] was not significant). Within the [+Ex] types, while the ratings for *that* and *if* did not differ from each other, they were both rated significantly higher than *when*, which was rated significantly higher than *because*.

Table 3. Summary of the mixed effects model for the raw ratings of the target sentences (adjunct islands). Significance indicators: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

	Estimate	SE	t	df	*p*	
that [–Ex] vs. *if* [–Ex]	0.333	0.120	2.774	143.403	0.006	**
if [–Ex] vs. *when* [–Ex]	−0.108	0.119	−0.907	136.577	0.366	
when [–Ex] vs. *because* [–Ex]	−0.293	0.118	−2.472	143.549	0.015	*
because [–Ex] vs. *that* [+Ex]	−0.826	0.138	−5.992	211.140	0.000	***
that [+Ex] vs. *if* [+Ex]	−0.043	0.130	−0.331	179.725	0.741	
if [+Ex] vs. *when* [+Ex]	−1.041	0.129	−8.083	179.660	0.000	***
when [+Ex] vs. *because* [+Ex]	−0.487	0.121	−4.042	136.090	0.000	***

4.2. Filler Type 1: Subject Islands

As in the target condition, the model had acceptability as outcome variable and type as predictor (with sliding contrasts). However, the only convergent model had no random slopes, only random intercepts for participant and item. The model is summarized in Table 4.

Table 4. Summary of the mixed effects model for filler type 1 (subject islands). Significance indicators: *** $p < 0.001$, marginal: · $p < 0.1$.

	Estimate	SE	t	df	*p*	
that [–Ex] vs. *if* [–Ex]	0.033	0.237	0.137	52.854	0.891	
if [–Ex] vs. *when* [–Ex]	0.134	0.237	0.565	52.854	0.575	
when [–Ex] vs. *because* [–Ex]	−0.225	0.237	−0.947	52.854	0.348	
because [–Ex] vs. *that* [+Ex]	−2.979	0.237	−12.549	52.854	0.000	***
that [+Ex] vs. *if* [+Ex]	0.046	0.237	0.192	52.854	0.849	
if [+Ex] vs. *when* [+Ex]	−0.421	0.237	−1.773	52.854	0.082	·
when [+Ex] vs. *because* [+Ex]	−0.269	0.237	−1.134	52.854	0.262	

The results and sliding pairwise comparisons for the eight sentence types are also shown in Figure 1B. There were no significant differences between the four baseline [−Ex] types, which were all rated in the top range. In contrast, the four [+Ex] sentence types were all in the lower range, and they were all significantly different from the [−Ex] types. In other words, there was a clear and significant negative main effect of extraction. There was also a marginally significant difference between *if* [+Ex] and *when* [+Ex] ($p = 0.082$).

4.3. Filler Type 2: Coordinate Structure

The converging model contained acceptability as outcome variable and type as predictor (with sliding contrasts) and random intercepts for participant and item, and random slopes for trial by participant and item. The model is summarized in Table 5.

Table 5. Summary of the mixed effects model for filler type 2 (coordinate structure). Significance indicators: *** $p<0.001$, ** $p < 0.01$, marginal: · $p < 0.1$.

	Estimate	SE	t	df	*p*	
that [–Ex] vs. *if* [–Ex]	0.268	0.150	1.790	54.999	0.079	·
if [–Ex] vs. *when* [–Ex]	0.052	0.150	0.347	52.842	0.730	
when [–Ex] vs. *because* [–Ex]	−0.430	0.143	−3.003	47.354	0.004	**
because [–Ex] vs. *that* [+Ex]	−3.518	0.142	−24.805	43.364	0.000	***
that [+Ex] vs. *if* [+Ex]	−0.104	0.136	−0.762	37.712	0.451	
if [+Ex] vs. *when* [+Ex]	0.073	0.130	0.560	34.270	0.579	
when [+Ex] vs. *because* [+Ex]	−0.162	0.128	−1.263	30.727	0.216	

See also Figure 1C for the results and sliding pairwise comparisons for the eight sentence types. Similar to the findings for filler 1, the acceptability ratings for the four baseline conditions [−Ex] were all in the top range, whereas the four [+Ex] sentence types were all in the lower range. There was a marginal difference between *that* [−Ex] and *if* [−Ex] and a significant difference between *when* [−Ex] and *because* [−Ex]. As with filler 1, the most dramatic effect is the negative main effect of extraction, as indicated by the difference between *because* [−Ex] and *that* [+Ex] ($p < 0.001$).

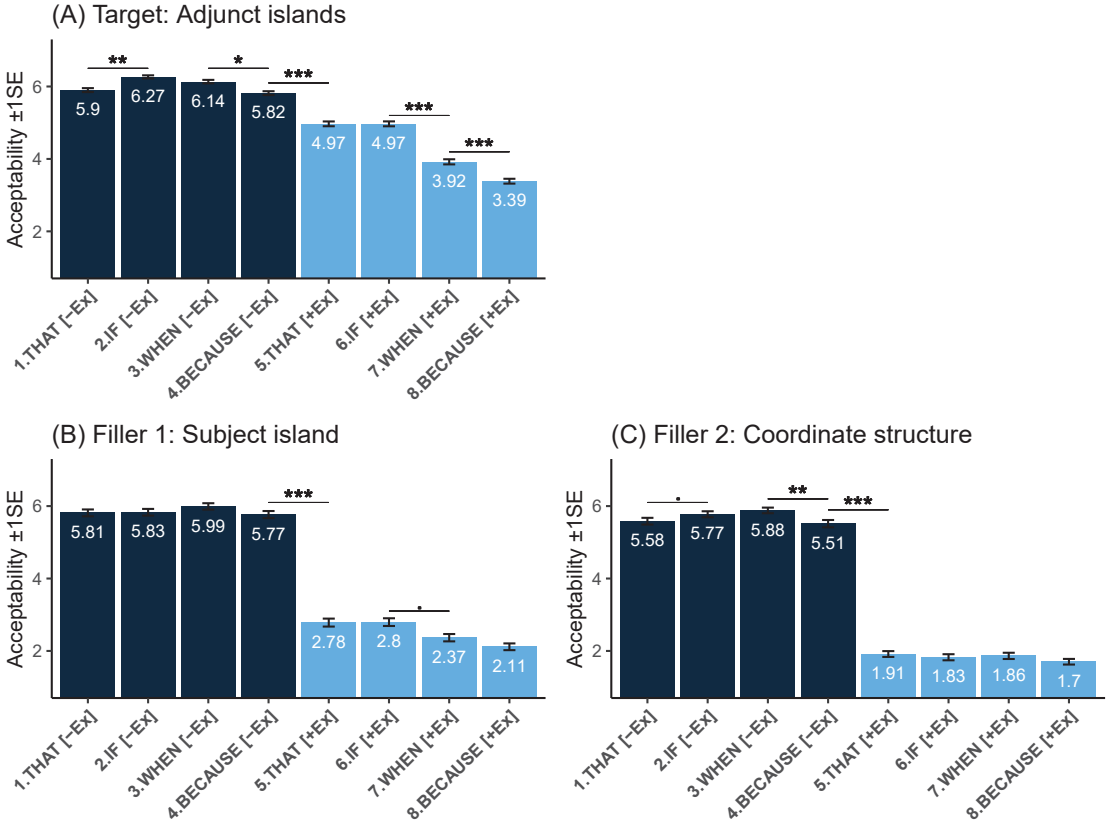


Figure 1. Mean (raw) acceptability ratings. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, marginal: · $p < 0.1$.

To see if the general acceptability pattern in the target condition (Figure 1A) was consistent across the 23 sets (or contexts, see Section 2), we plotted acceptability by type and participant for each set. As can be seen in Figure 2, the overall patterns look quite similar.

Next, in order to obtain an impression of the between-participant variation, we plotted the mean responses from each participant for the target sentences, see Figure 3. Although most of them show the same pattern with a downward slope from left to right, there is also a lot of variation. To control for this inter-participant variation (which is actually also taken into consideration as a random effect in the mixed-effects model) and to inspect it further, we applied *z*-transformation (next section).

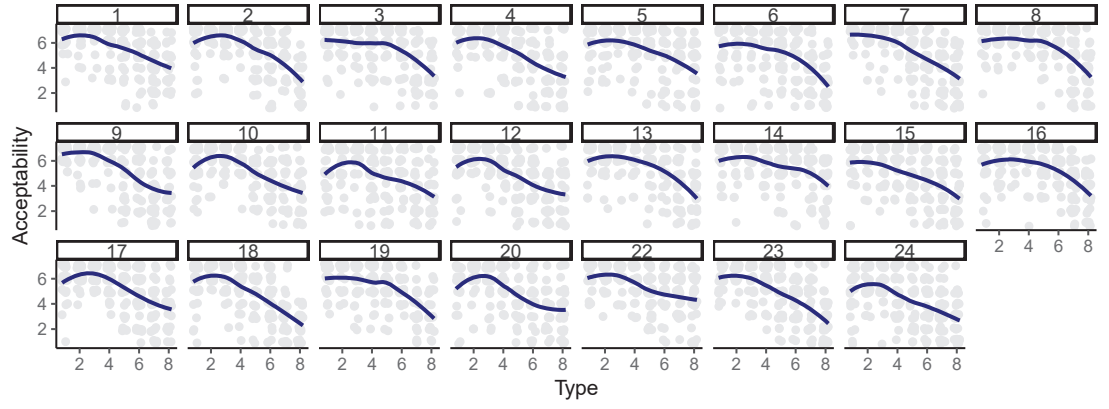


Figure 2. Mean (raw) acceptability ratings by type (see Figure 2) for each of the 24 target sets.

Figure 3. Mean (raw) acceptability ratings by type for each of the 235 participants.

4.4. Standardized (z-Transformed) Acceptability Ratings

For direct comparison with Sprouse et al. (2012, 2016) and a number of other studies, we *z*-score transformed the ratings by participant to control for potential individual scale bias, such as using only one end of the scale or a larger or smaller range (see also Kush et al. 2018, 2019; Bondevik et al. 2020). A *z*-transformed rating represents the number of standard deviations the raw (non-transformed) rating is from that participant's mean rating. As with the raw scores, the z-transformed results were analyzed using a linear mixed effects model with sliding contrasts. The model contained type as a predictor and random intercepts for participant and item, and random slopes for type and trial by participant, and random slopes for trial by item. The summary of the results is provided in Table 6 and plotted in Figure 4A. Note that the general acceptability pattern as well as the significant contrasts are the same as for the non-transformed ratings, compare Figures 1A and 4A, and Tables 3 and 6.

Table 6. Summary of the mixed effects model with sliding contrasts for the *z*-transformed ratings of the target sentences (adjunct islands). Significance indicators: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

	Estimate	SE	t	df	*p*	
that [–Ex] vs. *if* [–Ex]	0.172	0.060	2.858	148.388	0.005	**
if [–Ex] vs. *when* [–Ex]	−0.061	0.060	−1.015	141.736	0.312	
when [–Ex] vs. *because* [–Ex]	−0.155	0.060	−2.581	150.968	0.011	*
because [–Ex] vs. *that* [+Ex]	−0.377	0.067	−5.620	205.117	0.000	***
that [+Ex] vs. *if* [+Ex]	0.011	0.064	0.167	178.378	0.868	
if [+Ex] vs. *when* [+Ex]	−0.536	0.064	−8.400	180.962	0.000	***
when [+Ex] vs. *because* [+Ex]	−0.248	0.061	−4.084	151.319	0.000	***

To test for a superadditive 'island effect', we calculated the differences-in-differences (DD) score for each [+Ex] condition by subtracting the D score for *that*, measured as the difference in rating between *that* [–Ex] and *that* [+Ex] (i.e., the extraction effect for the non-island type), from the D scores for the three adjunct island types: *if*, *when*, and *because*, see Figure 4C,D (Sprouse et al. 2012; Kush et al. 2018, 2019; Bondevik et al. 2020). DD score proves a standardized measure of the island effect size which can be used for comparisons between different island types across experiments and languages. According to Kush et al. (2019, p. 401), "DD scores for island effects typically fall within the range of 0.75–1.25", and "any intermediate-sized island effect bears closer scrutiny". To test for significance, we analyzed the *z*-transformed ratings again using a linear mixed-effects model, this time with fixed effects of Complementizer and Extraction and their interaction, and random intercepts for participant and item, and random slopes for type and trial by participant, and random slopes for trial by item, first with *that* as intercept (summarized in Table 7), then with *when* as intercept (summarized in Table 8). (For *when*, the model had to be simplified such that it included random slopes only for extraction by participant).

Table 7. Summary of the mixed effects model of Complementizer and Extraction (*z*-transformed ratings, target sentences (adjunct islands)), using *that* as intercept. *** $p < 0.001$, ** $p < 0.01$, marginal: · $p < 0.1$.

	Estimate	SE	t	df	*p*	
Comp = *that* (Intercept)	0.598	0.047	12.787	181.579	0.000	***
Comp = *if*	0.190	0.066	2.895	176.200	0.004	**
Comp = *when*	0.118	0.066	1.787	182.055	0.076	·
Comp = *because*	−0.050	0.066	−0.746	185.248	0.457	
Extraction	−0.435	0.065	−6.702	171.523	0.000	***
Comp = *if* x Extraction	−0.162	0.092	−1.760	171.650	0.080	·
Comp = *when* x Extraction	−0.626	0.092	−6.812	171.573	0.000	***
Comp = *because* x Extraction	−0.724	0.092	−7.886	171.364	0.000	***

Table 8. Summary of the mixed effects model of Complementizer and Extraction (*z*-transformed ratings, target sentences (adjunct islands)), using *when* as intercept. *** $p < 0.001$, * $p < 0.05$, marginal: · $p < 0.1$.

	Estimate	SE	t	df	*p*	
Comp = *when* (Intercept)	0.716	0.046	15.456	175.595	0.000	***
Comp = *if*	0.071	0.066	1.087	175.994	0.279	
Comp = *because*	−0.168	0.065	−2.565	175.022	0.011	*
Comp = *that*	−0.119	0.065	−1.813	175.326	0.072	·
Extraction	−1.061	0.065	−16.212	175.326	0.000	***
Comp = *if* x Extraction	0.464	0.093	5.016	175.413	0.000	***
Comp = *because* x Extraction	−0.099	0.093	−1.068	175.109	0.287	
Comp = *that* x Extraction	0.626	0.093	6.759	175.227	0.000	***

Figure 4A shows the mean of the *z*-transformed acceptability ratings for the target condition, with and without extraction (compare Figure 1), 4B the extraction effects for the four complementizers, 4C the extraction effect sizes (all significant, $p < 001$, and *that*, *if*, and *when* significantly different form each other, $p < 0.001$), and 4D the DD scores indicating island effects. The DD score for *if* is only marginally significant (cf. Comp = *if* x Extraction, $p = 0.080$, in Table 7), whereas the DD scores for *when* and *because* were significant (Comp = *when* x Extraction and Comp = *because* x Extraction in Table 7, both $p < 0.001$). As indicated by Comp = *if* x Extraction, $p < 0.001$, in Table 8, the DD scores for *if* and *when* were significantly different from each other (and *when* and *because* were not, cf. Comp = *because* x Extraction, $p = 0.287$). Note that the DD scores are all smaller than 0.75, which, according to Kush et al. (2019, p. 401), is the lower threshold value for typical island effects.

In line with Kush et al. (2019) and Bondevik et al. (2020), in order to examine the underlying judgment pattern for each of the mean scores shown in Figure 1, we investigated the distribution of acceptability ratings by condition, as shown in the density plots in Figure 5. Uniform syntactic islands should show a unimodal distribution narrowly centered around $z = -1$, indicating that the participants consistently rated them as unacceptable. This is indeed the case for the two filler types with extraction (Figure 5B,C, light blue). The distribution of ratings for extractions from subject islands and coordinate structures is clearly skewed toward the left. Conversely, unambiguously well-formed types should show a unimodal distribution centered narrowly around $z = +1$, indicating that the participants consistently rated them as completely acceptable. Indeed, this is the case for all the baseline [−Ex] sentence types (Figure 5, dark blue): the ratings are clearly skewed towards the right. The same goes for extraction from complement clauses headed by *that*, though the distribution is more even and flat than the corresponding baselines (it is less pointy, suggesting more inter-participant variation). That leaves the three adjunct island extractions: *if*, *when*, and *because*. Extraction from *if* shows the same pattern as extraction from *that*, suggesting that extraction from *if* is equally acceptable (as shown in Figure 1 and Table 3). Extraction from *because* shows the inverse pattern, with a somewhat less pointy but clearly left-skewed unimodal distribution suggestive of islandhood. Finally, *when* stands out with a left-skewed but somewhat 'flatter' or even distribution, tending towards being bimodal with peaks at $z = -1$ and $z = 0$.

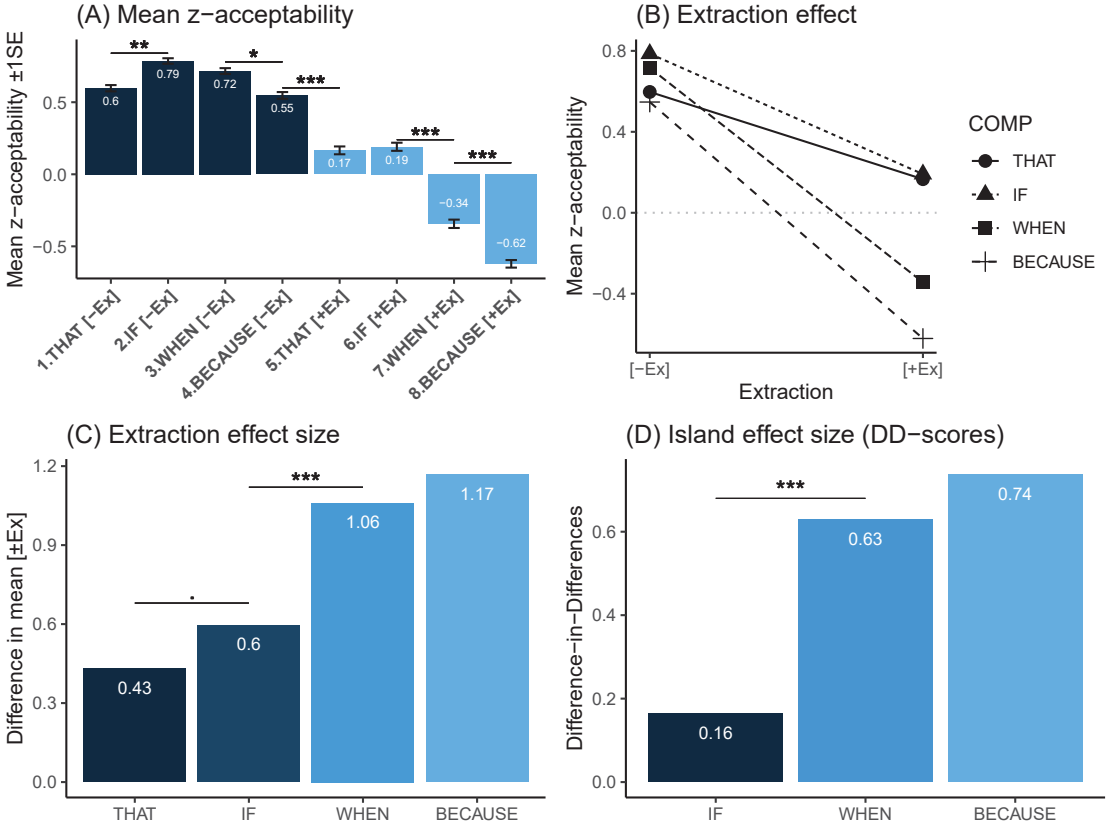


Figure 4. (**A**) Mean *z*-transformed acceptability ratings. (**B**) Extraction effect. (**C**) Extraction effect size (difference between [+Ex] and [–Ex]). (**D**) Island effect size measured in DD scores (Complementizer x Extraction interaction) between that and the other three complementizers). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, marginal: · $p < 0.1$.

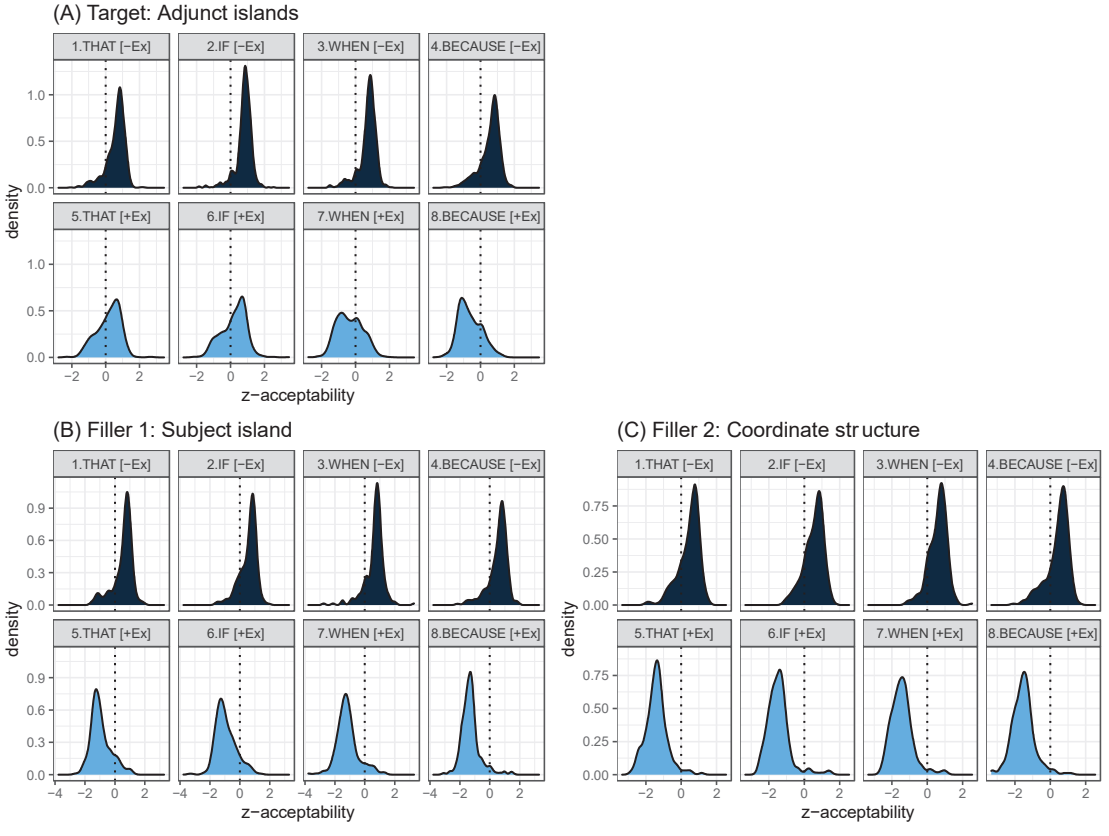


Figure 5. Density plots for each sentence type in the target category (**A**) and in the two fillers, (**B**) and (**C**). Dark blue: [−Ex], light blue: [+Ex].

4.5. Trial Effects

In order to test for trial effects, in particular positive correlations between repeated exposure of type and acceptability ratings, we plotted the mean acceptability (across participants on each list) of each occurrence of the four extraction types in the target category, see Figure 6A. *That*, *when*, and *because* showed a positive effect of trial, but the effect was only significant for *when*: in the course of the experiment, extraction across *when* increased 0.6 points on the 7-point scale ($R^2 = 0.17$, $p < 0.05$).

Under the (standard) assumption that adjunct clauses headed by *if*, *when*, and *because* have the same syntactic structure, we combined the three sets to see if their combined number of trials i.e., (9) would show a significant effect. This was not the case ($p > 0.1$, see Figure 6B).[4]

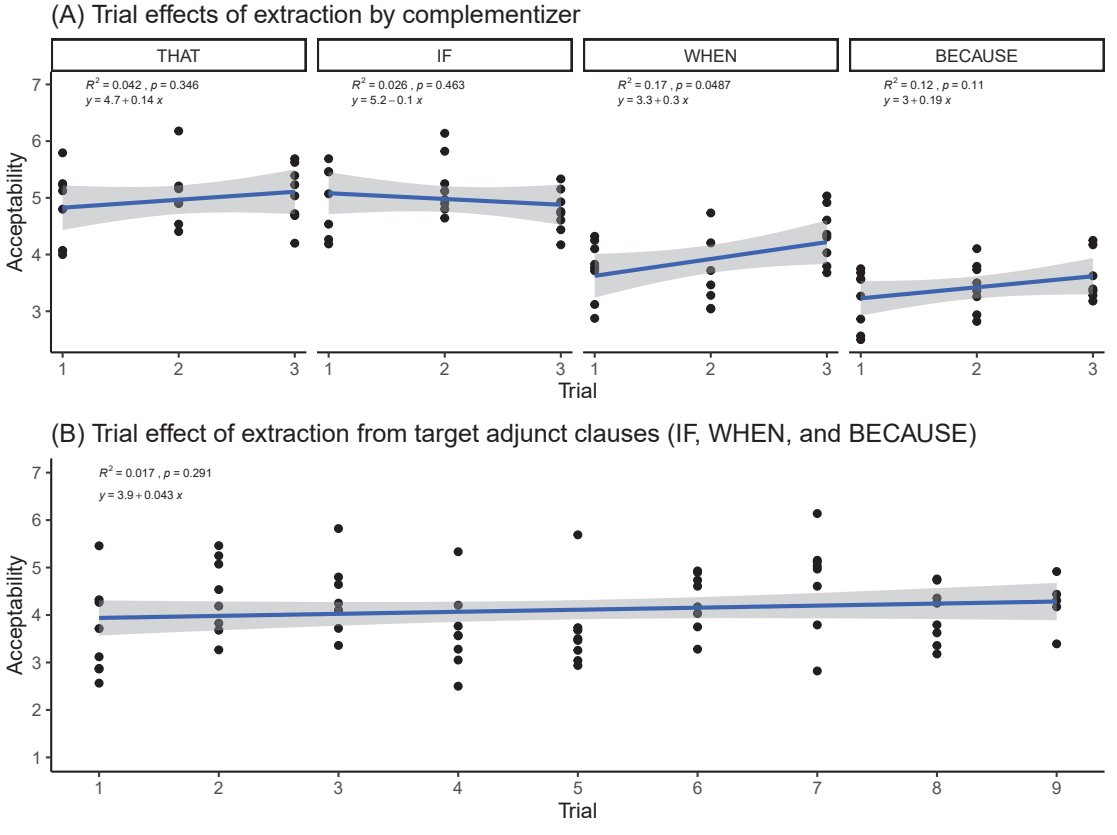


Figure 6. (**A**) Acceptability as a function of repetition (trial) of extractions from complement clauses headed by *that* and from adjunct clauses headed by *if*, *when* and *because* in the target category. Grey shading: 95% confidence interval. (**B**) Trial effect of extraction from target adjunct clauses headed by either *if*, *when* or *because*. Grey shading: 95% confidence interval.

5. Discussion

We investigated the acceptability of extraction from three different types of finite adjunct clauses (*if*-, *when*- and *because*-clauses) in English. Extraction was tested in the form of relativization (rather than *wh*-extraction) and in the presence of supporting context, since previous studies on island constructions in MSc. indicate that type of extraction dependency and contextual facilitation appear to have an impact on the possibility of at least some island extractions. Our study thus allowed us to examine whether finite adjunct clauses in English indeed constitute uniformly strong islands, as the traditional literature suggests, even when the potentially facilitating role of dependency type and context is taken into consideration. We predicted consistently low ratings for relativization from adjunct clauses, with no variation in acceptability across different adjunct clause types, and no trial effects. These predictions were not confirmed: the three different types of adjunct clauses tested in this study displayed a rather non-uniform behavior with regard to their acceptability. Specifically, extraction from *when*- and *because*-clauses both yielded significantly lower ratings than extraction from *if*-clauses. *If*-clauses appeared to pattern with non-island *that*-clauses instead, in that extraction from both *if*- and *that*-clauses yielded ratings above the middle range, and a similar, right-skewed distribution of acceptability ratings (as can be seen in the density plots in Figure 5). These findings suggest that, at

least for relativization, *if*-adjuncts are not absolute islands in English. However, *when*- and *because*-clauses in turn yielded not only significantly different ratings with extraction, but these structures also resulted in somewhat different acceptability distributions (with *because* showing a left-skewed distribution, whereas extraction from *when* resulted in a flatter distribution, cf. Figure 5). This picture is corroborated by a difference in the effects of repeated exposure: only *when*-clause extraction yielded a significant trial effect in our experiment, whereas the other clause types did not. *When*-clause extraction thus seems to stand out, both in terms of showing a trial effect and in showing a rather flat, somewhat bimodal distribution of acceptability ratings. This suggests that the reduced acceptability levels of the three clause types have different underlying causes (see also Snyder 2022).

Our finding that extraction from non-island *that*-clauses and from *if*-clauses both yielded positive and non-distinct ratings suggests that at least for relativization, *if*-clauses behave similarly to *that*-clauses in English in that they do not seem to universally block extraction. This finding challenges the strong island status of *if*-adjuncts. On the other hand, there was a statistically significant difference between *that* [−Ex] and *if* [−Ex], indicating that the baseline conditions for *that*- and *if*-clauses may not have been as minimally different as we intended. Some of the comments left by the participants in Google Forms reveal that a few participants were bothered by *that* occurring twice in relatively close proximity in our *that*-condition, as in (8). Both instances are actually optional, and some participants would prefer the sentences in which (at least) the first *that* was left out.

8. It's obvious **that** I was surprised **that** she actually completed this exercise.

The comments furthermore indicate that in a few items, the context provided by us may have been more felicitous with the conditional structure used in the *if*-clauses than with the indicative *that*-clause. Both of these circumstances could explain why our baseline *that*-clauses were rated significantly worse than the baseline *if*-clauses. Nevertheless, our conclusion that *if*-clauses do not appear to be absolute islands for relativization in English is supported by our finding that extraction from *if*-clauses yielded acceptability ratings that were clearly above the middle range and showed a right-skewed distribution of ratings, characteristic of acceptable sentences. These findings echo those of Kush et al. (2019), who found that the acceptability level of structures involving topicalization out of a conditional adjunct in Norwegian was on a par with the acceptability of topicalizing from a non-island embedded clause.

In the following sections, we discuss our findings in relation to dependency type (Section 5.1), the role of semantics and pragmatics in extractability (Section 5.2), processing factors (Section 5.3), and variation between complementizers and the internal structure of adjunct clauses (Section 5.4).

5.1. Extraction and Dependency Type

Our finding that adjuncts are not uniformly strong islands might similarly call for more fine-grained classifications or for a differentiated approach not only within the group of constructions traditionally referred to as adjunct islands, but also concerning the effect of dependency type on extractability. The fact that extraction from an *if*-clause appears to be rather acceptable in English may tell us something quite significant not only about the importance of applying a suitable context in this type of experiment, but also about the potential role of the type of dependency that links the filler and the gap. Previous experiments investigating extraction from finite *if*-clauses in English have primarily tested *wh*-extraction, and consistently yielded very low ratings for this type of extraction (e.g., Sprouse et al. 2012; Sprouse et al. 2016). A possible reason for the relatively high acceptability of extraction from *if*-adjuncts in our study is that extraction was in the form of relativization rather than question formation, and that extraction out of adjuncts may be possible for some dependencies such as relativization, but not for *wh*-movement. If this assumption is on the right track, the results from our study contribute to a growing body of evidence that extraction possibilities seem to differ across dependency types (for English, see Sprouse et al. 2016; Abeillé et al. 2020; for Norwegian, see Kush et al. 2019).

This difference in extraction dependencies is unexpected under traditional syntactic accounts of adjunct islands such as the CED: since *wh*-movement, topicalization and relativization are all instances of A'-dependencies, they should adhere to the same syntactic locality conditions. However, a syntactic analysis of this phenomenon may nevertheless be possible under the framework of *Relativized Minimality* (RM; Rizzi 1990), specifically under newer incarnations of RM called *featural RM* (e.g., Starke 2001; Rizzi 2013; Villata et al. 2016). Featural RM accounts for island effects in terms of feature-based intervention effects: a constituent cannot be extracted if the movement path crosses an intervening element that has the same featural specifications as the moved element. However, intervention effects can be overcome if the extracted phrase is more richly specified in morphosyntactic features than the potential intervener. An intervention account along these lines could potentially be applied to adjunct islands, in particular in the light of previous suggestions that some adjunct clauses (including conditional [*if*] and temporal clauses [*when*]) involve movement of an operator to the left periphery of the adjunct clause (e.g., Demirdache and Uribe-Etxebarria 2004; Bhatt and Pancheva 2006; Haegeman 2010, 2012). A potential asymmetry between different extraction dependencies into adjunct clauses could thus be accounted for under the assumption that the operator involved in the adjunct clause in question creates an intervention effect for *wh*-movement, but not for relativization across it. Such an account presupposes that relativization differs from *wh*-extraction in involving movement of an element that carries a richer feature specification than the operator present in the relevant adjunct clause (see also Sprouse et al. 2016 for a proposal along those lines).

A problem that this type of analysis faces is that it is not obvious how the featural setup of the relative operator (in relativization) or of the *wh*-phrase (in *wh*-movement) interacts with the type of operators that have been suggested to move in adjunct clauses in order to create (or overcome) intervention effects. Assuming that the proposed operators in *if*- and *when*-adjunct clauses differ in their feature specifications, it might provide a partial explanation of their different acceptability ratings. Crucially, however, in order to account for the gradience across all three types of adjuncts in our experiment, we would be forced to stipulate some kind of operator in *because*-clauses.

5.2. The Role of Semantics and Pragmatics in Extractability

Gradient acceptability may, however, not be surprising under a discourse-based account, such as the ones proposed recently by e.g., Chaves and Putnam (2020) (for a range of island phenomena) and Abeillé et al. (2020) (for subject islands specifically), echoing the longstanding idea that backgroundedness (Goldberg 2005) or pragmatic dominance (Erteschik-Shir 1973) are determining factors when it comes to extractability.

More specifically, Goldberg (2005, p. 135) argues that "Backgrounded Constructions are Islands" (BCI), which according to her definition means that it should be difficult to extract from a constituent that is neither primary topic nor part of the potential focus domain (Goldberg 2005, p. 130). Similarly, Abeillé et al. (2020) suggest a pragmatic account termed the *Focus-Background Conflict* (FBC) constraint, arguing that the island effects attested for certain constructions are due to the discourse clash that occurs when a focused element is part of a backgrounded constituent. They specifically apply this constraint to subject islands, where it can account for the relative acceptability of relativization from DP subjects in English and French, compared to *wh*-extraction from the same constituents: *Wh*-movement, but not relativization, puts the fronted element into focus, which clashes with the backgrounded nature of subjects. (Note, however, the very low scores for relativization from subjects in our experiment, cf. Figure 1B). An approach of this nature could be extended to cover the apparent possibility of relativizing from some adjunct clauses that seem to resist *wh*-extraction, given that adjunct clauses also typically express backgrounded information.

The three adjunct clause types in our stimuli are not discourse-functionally equivalent but have different semantico-pragmatic profiles. *If*, *when*, and *because* all introduce backgrounded adjunct clauses in the sense that the filler refers to a single member of a set

in the context, but they differ in terms of veridicality and presupposition: *when*- and *because*-clauses are veridical (they have positive polarity and a positive truth value), whereas *if*-clauses are not (they cannot be assigned a truth value). Furthermore, *when*-clauses and *because*-clauses (at least central ones) are presupposed, *if*-clauses are not (as they are not veridical).

What our data ultimately suggest is that the concept of *gradience* needs to be properly integrated into an account of island phenomena. Given that the acceptability of extracting from *if*-, *when*-, and *because*-clauses exhibits surprising variation, we still need to identify what exactly determines the differences in the acceptability of extraction from the adjunct clauses under investigation. According to Goldberg (2013, pp. 224–25), "backgrounded constituents of a sentence are not part of what is asserted by the sentence", and backgroundedness is gradient and independent of truth value, as being true is "not a requirement". In other words, to some extent, 'backgrounded' means 'not asserted'. One possible way of explaining the variation in the data while maintaining the basic assumptions of the BCI or the FBC could be to appeal to the idea of gradience in backgroundedness: the *because*-clauses could be argued to be somehow more 'asserted' than the *when*-clauses which in turn could be said to be more 'asserted' than the *if*-clauses (where *asserted* means ascribed a truth value). In that case, the clash between the adjunct clause and the filler-gap dependency relation by way of relativization should be greater for the *because*-clauses than e.g., the *if*-clauses, and this could give us the differences in acceptability found in the present study. It might be argued, then, that the more asserted a clause is, the less acceptable the extraction is predicted to be. *If*-clauses are not asserted, and they have no truth-value, and therefore, they would be predicted to be permeable and allow for extraction. However, a problem is, then, that the extractions from *when*-adjunct clauses in our experiment are much more acceptable than expected under this assumption; in a sense, our *when*-clauses are too 'good' (acceptable) to be 'true' (asserted), while our *because*-clauses are too 'true' (asserted) to be 'good' (acceptable). However, it is not clear that the relevant factor distinguishing between the three adjunct clause types is necessarily degree of assertion and/or backgroundedness (and, in any case, both terms still need to be properly defined and operationalized in terms of scalability). Other contenders might be degree of factivity (Kiparsky and Kiparsky 1970) or projectivity (Tonhauser et al. 2018)[5], realis/irrealis (Elliott 2000; Palmer 2001), veridicality (Giannakidou 1998), or indeed relevance, which we turn to now.[6]

As mentioned above, Chaves and Putnam (2020) argue that many island constraints traditionally assumed to be of a syntactic nature (e.g., the CED) can be reduced to "Relevance Islands": since adjuncts typically convey backgrounded information, extraction from them tends to violate a pragmatic principle requiring the referent that is singled out by the extraction to be sufficiently relevant for the main action, event, or state of affairs (as originally observed for extractions from complex subjects by (Kluender 2004). However, this pragmatic constraint can, according to Chaves and Putnam (2020, p. 91), be circumvented if extraction occurs in the form of relativization, since relatives "express assertions rather than backgrounded information". Chaves and King (2019) found a "positive correlation between the plausibility of the proposition (as expressed by a declarative clause) and the extractability of a deeply embedded referent therein" (Chaves and Putnam 2020, p. 222). The difference in acceptability between the three types of adjunct island complementizers found in the present study might thus be related to the proposition that they create. It seems reasonable to assume that the semantic plausibility of a proposition plays a role in connection with extractability. It is, of course, important that the corresponding declarative or non-extracted version of a sentence be plausible and felicitous (i.e., coherent) in order for extraction to be acceptable (see also Engdahl 1997; Kehl 2022). However, more specifically, the more prototypical and coherent the relations between the "semantic components" are, the easier it is to conceive of a referent as relevant to the main event and hence to extract it (Chaves and Putnam 2020, pp. 222–23). This approach may account for some of the variability attested at the participant level, given that the likelihood of an extracted referent being considered relevant to the main event by the comprehender may be a matter of degree and

thus subject to between-participant variation. However, our results suggest rather uniform interpretations across participants, except for extraction from adjunct clauses headed by *when* (see Figures 3 and 5).

Furthermore, the account presented in Chaves and Putnam (2020) suffers from the same problems that other semantico-pragmatic accounts have, namely the difficulty in operationalizing the approach in a falsifiable way. In fact, one of the few tools that they do suggest makes only partially correct predictions. Specifically, Chaves and Putnam (2020, p. 235) suggest that the acceptability of extracting from an adjunct clause is correlated with the inter-clausal semantic relations hierarchy in Van Valin (2005, pp. 208–9), according to which causal relations express a tighter semantic link or a higher degree of cohesion than purely temporal relations. However, as mentioned above, all three adjunct clause types in our study were embedded under psych-predicates that trigger a causal reading of the events in the matrix and embedded clause. If a causal reading is equally available for all three types of adjunct clauses in our study, they should rank equally high on Van Valin's hierarchy of semantic relations and also allow extraction equally easily, which is not what we found. Possibly, *when*-clauses can be argued to express a smaller degree of cohesion than *if*- and *because*-clauses since the causal relation is not explicitly encoded in *when*-clauses but must be inferred pragmatically. *If*- and *because*-clauses, on the other hand, linguistically encode causation by means of the introducing element (*if* and *because*). This would explain why extraction from *when*-clauses scored significantly lower ratings than extraction from *if*-clauses in our study. However, it leaves unexplained why extraction from *because*-clauses was rated significantly worse than both *if*- and *when*-clause extraction.[7]

5.3. Processing Factors

When- and *if*-clauses may, however, differ from *because*-clauses in terms of processing, given that they can occur as a complement or argument. In examples such as (9a) and (9b), the matrix verb *appreciate* appears to select *if*- and *when*-clauses which hence function as arguments in the superordinate clause. Crucially, however, this is not possible with *because*-clauses, (9b) and (10b) (see Teleman et al. 1999, pp. 568–93 for a similar observation in relation to Swedish). (Examples (9a) and (10a) are from COCA).

9. a. I would appreciate if you can post a link to my article to everyone in your network
   b. *I appreciate because you can post a link to my article to everyone in your network.
10. a. Sir, we always appreciate when you come on.
    b. *Sir, we always appreciate because you come on.

Based on these data, it seems possible that speakers may be able to assign a structural parse to some sentences involving *if*- and *when*-clauses where the *if*- or *when*-clause has the status of an argument or complement (similar to the *that*-clauses used in our stimuli), whereas this is not an option for *because*-clauses. While the selectional properties of predicative adjectives are different from those of transitive verbs such as *appreciate*, it may be the case that the parser, holding a filler, postulates an intermediate attachment site at the left edge of the adjunct clause, for example via the principle of *Attach Anyway* (Fodor and Inoue 1998). This phenomenon is reminiscent of the effects of matrix verb compatibility found for long-distance dependencies: when it is plausible that the filler originates in the matrix clause and intermediate attachment is possible, the resulting acceptability level is higher than if the matrix verb is not compatible with the filler (see Christensen et al. 2013 and references cited there). This in turn lends credence to the CED: while it seems nearly impossible to salvage a purely syntactic version of CED, the latter may be couched in processing or computational terms instead. It is simply easier to parse extractions from complements than adjuncts, and the data in our study may at least partially be explained as a misparse of the adjunct clauses as complement clauses, thus a local ambiguity leading to a global (positive) effect on acceptability. The variation found between *if*, *when* and *because* might be due to the frequency with which they occur as complement clauses (for instance,

COCA returns 2.130 hits for "appreciate that" with *that* specified as a complementizer, 200 for "appreciate if" and 103 for "appreciate when", while there are no hits of the relevant type for "appreciate because"). The possibility of extracting from adjunct clauses may thus be modulated by frequency, as the parser may use probabilistic knowledge in the online processing of the structures involving extraction from adjunct clauses in the sense that it may temporarily postulate intermediate attachment of *if*- and *when*-clauses as complements rather than adjuncts, based on previous exposure, and perhaps this may ameliorate the acceptability level overall. Future research, e.g., using G-Maze (Witzel et al. 2012), could investigate the online processing of these structures and whether the comprehender has a slower reaction time at the complementizer *because* than for *if* and *when*, which would lend support to the idea that the syntactic integration is more difficult, potentially resulting in a decrease in acceptability.

On the other hand, the potential amelioration of temporarily interpreting adjuncts as complements would be the opposite of what has been observed for garden-path resolution. Studies on parsing and ambiguity resolution has shown that garden-path sentences are particularly difficult because the position where the ambiguous constituent is initially integrated does not contain the (correct) target position (Pritchett 1992). In a (non-garden path) sentence such as *She knew Tom liked Cats*, *Tom* is temporarily attached as direct object of *knew*, but has to be reanalyzed as the subject of the embedded clause in the object position. In the garden-path sentence *While she dressed Tom made a sandwich*, *Tom* is again first analyzed as object, but has to be reanalyzed as the subject of the matrix clause in which *while she dressed* is embedded. This reanalysis is effortful and costly, leading to a severe reduction in acceptability.

5.4. Variation between Complementizers and the Internal Structure of Adjunct Clauses

As discussed in Section 5.1, the variation in acceptability as a function of the choice of complementizer (*if*, *when*, *because*) is difficult to explain under a purely syntactic account (or solely under a processing account, see Section 5.3) given that the underlying structure of the adjunct island constructions is generally assumed to be basically the same. However, some have argued that different types of adjunct clauses are adjoined at different positions. For instance, Haegeman (2012) distinguishes between *central* and *peripheral adverbial clauses*, which contrast in their relation to the clause that they modify, as the central adverbial clauses are event structuring (modifying the matrix clause at the event level), while the peripheral ones are discourse structuring. In other words, central adjunct clauses modify the event encoded in the host proposition, while peripheral adjunct clauses provide a contextually accessible background assumption contributing evidence for the relevance of the matrix domain (Badan and Haegeman 2022). They differ in attachment height, such that peripheral adjuncts are adjoined to CP, central ones to TP or *v*P (Haegeman 2012, p. 171), and correspondingly, in their degree of opacity for syntactic operations. However, all three types of adjunct clauses used in this study (*if*-, *when*- and *because*-adjuncts) are central adjunct clauses that are merged at the same height according to Haegeman's (2012, p. 164) classification. Other proposals connect differences in attachment height to differences in whether or not the adjunct can be interpreted as being in a coherent (e.g., causal) single-event relation with the matrix clause (e.g., Ernst 2001; Narita 2011; Truswell 2011; Sheehan 2013; Brown 2017). However, since all of our target sentences were construed to trigger a semantically coherent interpretation between the matrix and adjunct clause events, these proposals cannot capture the differences across different adjunct clauses observed in our study.

There are some indications that causal adjuncts (*because*-clauses) have more elaborate syntactic structures than conditional (*if*-clauses) and temporal adjuncts (*when*-clauses). In terms of Haegeman's (2012) classification, causal clauses are merged at the same height as other central adjunct clauses, but seem to display the internal syntax of peripheral adverbial clauses (Müller 2019, p. 98). For example, causal clauses can contain epistemic modal markers, which are usually only possible in peripheral clauses (Ros 2005, p. 98).

Causal clauses are also compatible with V2 word order in the MSc. languages, see Teleman (1967) for Swedish, Julien (2015) for Norwegian, and Nyvad et al. (2017, p. 462) for Danish. Based on this evidence, it has been suggested that causal clauses are root-like in that they can encode independent illocutionary force (Müller 2019), which is also in line with the observation that causal clauses may be asserted independently (Hooper and Thompson 1973; Sawada and Larson 2004), whereas temporal and conditional clauses are considered non-assertive (see also Section 5.2).

An account that may be compatible with the presented facts is the *c*P/CP-distinction first proposed by Nyvad et al. (2017) for Danish: these authors propose an account which credits evident island extractability in Danish (and presumably the other Mainland Scandinavian languages) to the possibility of CP-recursion. Given the apparent acceptability of the structures in question, the grammar must be able to generate them (i.e., the syntax must provide a licit hierarchical representation), and given that cross-clausal filler-gap dependencies adhere to the principle of successive-cyclic movement (e.g., Chomsky 1973; 1995, 2001; Den Dikken 2009; van Urk 2020; Davis 2021), an 'escape hatch' must be present. The latter may be provided by *c*P ("little *c*P"), a projection found in embedded clauses, which hosts functional elements such as complementizers and which can also be recursive, thus accounting for both the widespread use of complementizer stacking and otherwise unexplained island extractability in e.g., Danish. "Big" CP, on the other hand, hosts lexical elements such as finite verbs in (embedded and main clause) V2. Under this analysis, CPs are always finite, they carry illocutionary force (or illocutionary potential) and they are strong islands. Hence, in this framework, CPs are root clauses or root-like clauses. A similar observation is made in connection with Haegeman's (2002) proposal that (root-like) peripheral adjunct clauses may have a Force projection, allowing embedded topicalization and V2, while central adjunct clauses do not. See Haegeman (2012, p. 155) for examples of the impossibility of argument fronting in central *if*- and *when*-clauses in English.

It may hence be the case that significantly decreased acceptability found for extraction from *because*-clauses (as compared to *if*- and *when*-clauses) is due to the internal syntax of the *because*-clauses being more complex compared to *if*- and *when*-clauses. Whether that is the result of the *because*-clauses having a Force projection, a root-like C-projection or simply a feature encoding illocutionary force is still an open question, as is the question of how similar English really is to the Mainland Scandinavian languages when it comes to island extractability: while the Mainland Scandinavian languages appear to be able to extract fillers from embedded questions (e.g., Christensen et al. 2013), relative clauses (e.g., Lindahl 2017) and adjunct clauses (e.g., Müller 2019) more or less independently of dependency type (i.e., topicalization, *wh*-movement or relativization), English may be more limited in its range of possibilities.

Overall, however, relativization out of adjunct clauses appears to be possible in English when presented with a supporting context. *If*-, *when*- and *because*-clauses may thus not be strong islands after all, and they do not form a uniform class, as they seem to have different profiles in our data: the acceptability of extraction from *if*-clauses was not significantly different from that of *that*-clauses, even though the two types of structures are viewed as islands (adjuncts) and non-islands (complements), respectively. With regards to the *when*-clauses, they stand out in a different way: the density plot (acceptability distribution) for this clause type is flatter and slightly bimodal, indicating more variation between participants. In addition, there was a trial effect for the *when*-clauses such that the participants in the study found extraction significantly better over time, but not for *because*-clauses, which suggests different underlying causes for reductions in acceptability (Snyder 2022). Following Sprouse (2007, p. 124), this trial effect suggests that extraction from *when*-clauses is licit, since only grammatical structures should exhibit priming effects (see also Christensen et al. 2013; Christensen and Nyvad 2014). The acceptability of *because*-clauses did not improve significantly over time, but the DD-score for this structure type was actually below the threshold of 0.75 for islandhood proposed by Kush et al. (2019).

In short, none of the adjunct clause types showed clear signs of strong islandhood. The present study also illustrates the importance of applying different types of statistical analyses. Without this type of 'profiling', the multifactorial profiles of extraction from English adjunct clauses would go undetected. We suspect that the acceptability ratings result from interactions between subtle syntactic contrasts, pragmatics, and processing factors. It is very difficult, if not impossible, to tease apart the exact contribution of each of the individual factors to the observed pattern of acceptability.

6. Conclusions

Universal constraints should hold cross-linguistically and across constructions, and certainly across different syntactic constructions of the same underlying type. Apparent cross-linguistic variation in adjunct islands, as displayed by the MSc. languages, has posed a challenge for both formal syntactic accounts and processing accounts of island constraints. However, the reports of crosslinguistic variation are based on comparisons of rather different structures in English and Scandinavian (usually, non-contextualized extraction by *wh*-movement in English vs. contextually facilitated topicalization in MSc.). Once the impact of different extraction dependencies and of the presence of context on extractability is taken into consideration, the emerging picture is that the crosslinguistic variation between these languages regarding adjunct islandhood may have been exaggerated. Even though our experiment does not directly compare English to other languages, the results are compatible with the view that English is similar to MSc. in that different adjunct clauses display a non-uniform behavior when it comes to extraction. More specifically, the pattern observed in our study for different adjunct clauses matches the patterns found for topicalization from the corresponding adjuncts in Swedish (Müller 2017) and Norwegian (Bondevik 2018; Bondevik et al. 2020): extraction from Swedish and Norwegian *if*-clauses yielded smaller island effects and better average ratings than extraction from *when*-clauses, which in turn yielded smaller island effects and better ratings than extraction from *because*-clauses. This suggests that English and the MSc. languages might be more similar than previously assumed with regards to adjunct islandhood.

The heterogeneity in the acceptability patterns related to adjunct island violations suggests that syntax alone cannot account for the data. Syntactic theory may possibly need to be modified in the face of mounting evidence suggesting that configurational syntax alone cannot explain the variability in acceptability attested across a wide range of traditional island structures and that extra-grammatical factors are also essential in understanding these phenomena.

Author Contributions: Conceptualization, A.M.N. and K.R.C. Methodology, A.M.N., C.M. and K.R.C. Investigation, A.M.N., C.M. and K.R.C. Data curation, K.R.C. Formal analysis and visualization, K.R.C. (with contributions from A.M.N.). Writing—original draft, A.M.N., C.M. and K.R.C. (A.M.N. and C.M. mainly responsible for writing Sections 1, 2, 5 and 6, K.R.C. mainly responsible for writing Section 4). Writing—review and editing, A.M.N., C.M. and K.R.C. Project administration and funding acquisition, A.M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Independent Research Fund Denmark (grant ID: DFF-9062-00047B).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The stimulus set used in this study is openly available at the Open Science Framework: https://osf.io/yv43b/?view_only=bf33d6349d7a495d88758fc5a0d7822e (accessed on 4 September 2022).

Acknowledgments: We would like to thank three anonymous reviewers for their constructive criticism and helpful comments. We would also like to thank Clara Ulrich Eggers for tireless discussions on the stimulus sentences, and Christopher Martin for help with the stimuli.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Notes

1 PRT = particle.

2 We follow Kehler's (2002) classification of three classes of Coherence relations (*Resemblance*, *Cause-Effect*, and *Contiguity*). According to this classification, adjunct clauses with a purely temporal reading are less coherent than their causally interpreted counterparts, since they convey only one type of coherence relation between matrix and adjunct clause (i.e., Contiguity), whereas adjuncts with a causal reading convey both a Contiguity and a Cause-Effect relation to the matrix. Similarly, Jin (2015) and Van Valin and LaPolla (1997, p. 478) rank different coherence relations on a *cohesiveness scale*. Causal relations convey a higher degree of coherence on this scale than purely temporal relations.

3 Chaves and Putnam (2020, p. 72ff) argue that the relevant constraint here is the Conjunct Constraint, which blocks extraction *of* a conjunct, whereas extraction *from* a conjunct is blocked by the Element Constraint. Only the Conjunct Constraint is absolute. Here, we use the Coordinate Structure Constraint as synonymous with the Conjunct Constraint (see also Culicover et al. 2022, pp. 21–22).

4 The fact that we did not find a trial (or repeated exposure) effect for *if* and *because* may be argued to be due to the number of instantiations of adjunct islands that the participants were exposed to. Based on work by Hofmeister et al. (2015), Chaves and Putnam (2020, p. 229) argue that "Adjunct Island violations ameliorate only if the number of exposures is sufficiently high". They show that repeated exposure effects did not arise for participants exposed to six repetitions of adjunct island violations, whereas exposure to twelve instantiations were enough to demonstrate a positive trial effect over time (Chaves and Putnam 2020, pp. 232–33). In our stimuli, each participant was exposed to 9 adjunct island violations, 3 extractions from each type (*if*, *when*, and *because*).

5 The *Gradient Projection Principle* (Tonhauser et al. 2018, p. 499) states that "If content C is expressed by a constituent embedded under an entailment-canceling operator, then C projects to the extent that it is not at-issue." The term "projection" here refers to the extent to which the speaker "may be taken to be committed to" the content of C, i.e. the extent to which C is entailed. In our experiment, however, all the embedded clauses in our stimuli are at issue, because they are part of the context and causally linked to their matrix clause.

6 However, we have difficulty translating the observed gradience (*if* > *when* > *because*) into realis or veridicality without running into the same kind of problems. Only *if* is irrealis and non-veridical (neither veridical or antiveridical); only *when* is not explicitly causal (it has to be inferred), and hence, neither realis or veridicality can account for the observed gradience.

7 In a related vein, Truswell (2011, p. 157) proposes a semantic account of the possibility of adjunct clause extraction in English, namely the *Single Event Grouping Condition* (SEGC) which states that *wh*-dependencies into adjunct clauses are allowed "if the minimal constituent containing the head and the foot of the chain can be construed as describing a single event grouping". A single event grouping requires a spatiotemporal overlap between the matrix clause and the adjunct clause, as well as maximally one verb in the two clauses being agentive. In addition, Truswell (2011) argues that there is a finiteness operator in finite adjunct clauses, blocking extraction from such structures. However, as our results show, extraction from finite adjunct clauses in English is actually possible. While the SEGC would allow us to make distinctions between structurally parallel adjunct clauses, its predictions do not match our results, which show that extractions are possible to varying degrees. Furthermore, given that the truth of the most deeply embedded clause is a prerequisite for the emotional state expressed by the psych-predicate in the matrix clause, there is a causal connection that leads to a potential spatiotemporal overlap uniformly across the three adjunct types, and so the variation attested is unexpected.

References

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from Subjects: Differences in Acceptability Depend on the Discourse Function of the Construction. *Cognition* 204: 104293. [CrossRef] [PubMed]

Abrusán, Márta. 2014. *Weak Island Semantics*, 1st ed. Oxford Studies in Semantics and Pragmatics 3. Oxford: Oxford University Press.

Augie, Baptiste. 2016. Package 'GridExtra'. Available online: https://github.com/baptiste/gridextra (accessed on 13 September 2021).

Badan, Linda, and Liliane Haegeman. 2022. The Syntax of Peripheral Adverbial Clauses. *Journal of Linguistics*, 1–42. [CrossRef]

Bernardy, Jean-Philippe, Shalom Lappin, and Jey Han Lau. 2018. The Influence of Context on Sentence Acceptability Judgements. In *Proceedings of 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne: Association for Computational Linguistics, pp. 456–61. [CrossRef]

Bhatt, Rajesh, and Roumyana Pancheva. 2006. Conditionals. In *The Blackwell Companion to Syntax*. Edited by Martin Everaert and Henk van Riemsdijk. Malden: Blackwell Publishing, vol. 1, pp. 638–87. [CrossRef]

Bondevik, Ingrid. 2018. Investigating the Universality of Adjunct Islands through Formal Acceptability Experiments. Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway. Available online: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2573472 (accessed on 30 March 2022).

Bondevik, Ingrid, Dave Kush, and Terje Lohndal. 2020. Variation in Adjunct Islands: The Case of Norwegian. *Nordic Journal of Linguistics* 44: 223–54. [CrossRef]
Brown, Jessica M. M. 2017. Heads and Adjuncts: An Experimental Study of Subextraction from Participials and Coordination in English, German and Norwegian. Ph.D. dissertation, University of Cambridge, Cambridge, UK. [CrossRef]
Chaves, Rui P., and Adriana King. 2019. A Usage-Based Account of Subextraction Effects. *Cognitive Linguistics* 30: 719–50. [CrossRef]
Chaves, Rui P., and Michael T. Putnam. 2020. *Unbounded Dependency Constructions: Theoretical and Experimental Perspectives*. Oxford: Oxford University Press.
Chomsky, Noam. 1973. Conditions on Transformation. In *A Festschrift for Morris Halle*. Edited by Stephen R. Anderson and Paul Kiparsky. New York: Holt, Rinehart and Winston, pp. 232–86.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Studies in Generative Grammar 9. Dordrecht: Foris Publications.
Chomsky, Noam. 1995. *The Minimalist Program*. Current Studies in Linguistics 28. Cambridge: MIT Press.
Chomsky, Noam. 2001. Derivation by Phase. In *Ken Hale: A Life in Language*. Edited by Michael J. Kenstowicz. Cambridge: MIT Press, pp. 1–52.
Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2014. On the Nature of Escapable Relative Islands. *Nordic Journal of Linguistics* 37: 29–45. [CrossRef]
Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2019. No Escape from the Island: On Extraction from Complement Wh-Clauses in English. In *The Sign of the V—Papers in Honour of Sten Vikner*. Edited by Ken Ramshøj Christensen, Henrik Jørgensen and Johanna L. Wood. Aarhus: Dept. of English, School of Communication & Culture, Aarhus University, pp. 95–112. [CrossRef]
Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2022. The Island Is Still There: Experimental Evidence for The Inescapability of Relative Clauses in English. *Studia Linguistica*, 1–25. [CrossRef]
Christensen, Ken Ramshøj, Johannes Kizach, and Anne Mette Nyvad. 2013. Escape from the Island: Grammaticality and (Reduced) Acceptability of Wh-Island Violations in Danish. *Journal of Psycholinguistic Research* 42: 51–70. [CrossRef]
Culicover, Peter W., and Susanne Winkler. 2022. Parasitic Gaps Aren't Parasitic, or, the Case of the Uninvited Guest. *The Linguistic Review* 39: 1–35. [CrossRef]
Culicover, Peter W., Giuseppe Varaschin, and Susanne Winkler. 2022. The Radical Unacceptability Hypothesis: Accounting for Unacceptability without Universal Constraints. *Languages* 7: 96. [CrossRef]
Davies, Mark. 2008. The Corpus of Contemporary American English (COCA): 560 Million Words, 1990–Present. Available online: https://www.english-corpora.org/coca/ (accessed on 30 March 2022).
Davis, Colin P. 2021. Possessor Extraction in Colloquial English: Evidence for Successive Cyclicity and Cyclic Linearization. *Linguistic Inquiry* 52: 291–332. [CrossRef]
Deane, Paul. 1991. Limits to Attention: A Cognitive Theory of Island Phenomena. *Cognitive Linguistics* 2: 1–64. [CrossRef]
Demirdache, Hamida, and Myriam Uribe-Etxebarria. 2004. The Syntax of Time Adverbs. In *The Syntax of Time*. Edited by Jacqueline Guéron and Jacqueline Lecarme. Cambridge: MIT Press, pp. 143–80. [CrossRef]
Den Dikken, Marcel. 2009. Arguments for Successive-Cyclic Movement through SpecCP A Critical Review. *Linguistic Variation Yearbook* 9: 89–126. [CrossRef]
Fodor, Janet Dean, and Atsu Inoue. 1998. Attach Anyway. In *Reanalysis in Sentence Processing*. Edited by Janet Dean Fodor and Fernanda Ferreira. Dordrecht: Springer Netherlands, pp. 101–41. [CrossRef]
Elliott, Jennifer R. 2000. Realis and Irrealis: Forms and Concepts of the Grammaticalisation of Reality. *Linguistic Typology* 4: 55–90. [CrossRef]
Engdahl, Elisabet. 1997. Relative Clause Extraction in Context. *Working Papers in Scandinavian Syntax* 60: 59–86.
Engelkamp, Johannes, and Hubert D. Zimmer. 1983. *Dynamic Aspects of Language Processing: Focus and Presupposition*. Springer Series in Language and Communication 16; Berlin and Heidelberg: Springer. [CrossRef]
Ernst, Thomas. 2001. *The Syntax of Adjuncts*. Cambridge: Cambridge University Press. [CrossRef]
Erteschik-Shir, Nomi. 1973. On the Nature of Island Constraints. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge.
Erteschik-Shir, Nomi. 2006. Bridge Phenomena. In *The Blackwell Companion to Syntax*. Edited by Martin B. H. Everaert and Henk C. van Riemsdijk. Malden: Blackwell Publishing, pp. 284–94. [CrossRef]
Giannakidou, Anastasia. 1998. *Polarity Sensitivity as (Non)Veridical Dependency*. Amsterdam: John Benjamins. [CrossRef]
Goldberg, Adele E. 2005. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press. [CrossRef]
Goldberg, Adele E. 2013. Backgrounded Constituents Cannot Be 'Extracted'. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 221–38. [CrossRef]
Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory*, 2nd ed. Blackwell Textbooks in Linguistics 1. Oxford: Blackwell.
Haegeman, Liliane. 2002. Anchoring to Speaker, Adverbial Clauses and the Structure of CP. *Georgetown University Working Papers in Theoretical Linguistics* 2: 117–80.
Haegeman, Liliane. 2010. The Internal Syntax of Adverbial Clauses. *Lingua, Exploring the Left Periphery* 120: 628–48. [CrossRef]
Haegeman, Liliane. 2012. *Adverbial Clauses, Main Clause Phenomena, and the Composition of the Left Periphery: The Cartography of Syntactic Structures, Volume 8*. Oxford Studies in Comparative Syntax. New York: Oxford University Press. [CrossRef]
Hofmeister, Philip, Peter Culicover, and Susanne Winkler. 2015. Effects of Processing on the Acceptability of 'Frozen' Extraposed Constituents. *Syntax* 18: 464–83. [CrossRef]
Hooper, Joan B., and Sandra A. Thompson. 1973. On the Applicability of Root Transformations. *Linguistic Inquiry* 4: 465–97.

Huang, Cheng-Teh James. 1982. Logical Relations in Chinese and the Theory of Grammar. Ph.D. dissertation, MIT, Cambridge, UK. Available online: http://pubman.mpdl.mpg.de/pubman/item/escidoc:403106:7/component/escidoc:403105/chinese_huang1982_o.pdf (accessed on 30 March 2022).
Jin, Dawei. 2015. Coherence Relation and Clause Linkage: Towards a Discourse Approach to Adjunct Islands in Chinese. *Studies in Language. International Journal Sponsored by the Foundation Foundations of Language* 39: 424–58. [CrossRef]
Julien, Marit. 2015. The Force of V2 Revisited. *The Journal of Comparative Germanic Linguistics* 18: 139–81. [CrossRef]
Kehl, Andreas. 2022. Extraction from Present Participle Adjuncts: The Relevance of the Corresponding Declaratives. *Languages* 7: 177. [CrossRef]
Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford: CSLI Publications.
Kiparsky, Paul, and Carol Kiparsky. 1970. Fact. In *Progress in Linguistics: A Collection of Papers*. Edited by Manfred Bierwisch and Karl Erich Heidolph. Berlin: De Gruyter, pp. 143–73. [CrossRef]
Kluender, Robert. 1998. On the Distinction between Strong and Weak Islands: A Processing Perspective. In *The Limits of Syntax*. Edited by Peter W. Culicover and Louise McNally. San Diego: Academic Press, pp. 241–79.
Kluender, Robert. 2004. Are Subject Islands Subject to a Processing Account? In *Proceedings of WCCFL*. Somerville: Cascadilla Press, vol. 23, pp. 101–25.
Kristensen, Line Burholt, Elisabeth Engberg-Pedersen, and Mads Poulsen. 2014. Context Improves Comprehension of Fronted Objects. *Journal of Psycholinguistic Research* 43: 125–40. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating Variation in Island Effects: A Case Study of Norwegian Wh-Extraction. *Natural Language & Linguistic Theory* 36: 743–79. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the Island Sensitivity of Topicalization in Norwegian: An Experimental Investigation. *Language* 95: 393–420. [CrossRef]
Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. LmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82: 1–26. [CrossRef]
Lau, Jey Han, Carlos Armendariz, Shalom Lappin, Matthew Purver, and Chang Shu. 2020. How Furiously Can Colorless Green Ideas Sleep? Sentence Acceptability in Context. *Transactions of the Association for Computational Linguistics* 8: 296–310. [CrossRef]
Lindahl, Filippa. 2017. *Extraction from Relative Clauses in Swedish*. Gothenburg: University of Gothenburg. Available online: https://gupea.ub.gu.se/handle/2077/51985 (accessed on 30 March 2022).
Michel, Dan, and Grant Goodall. 2013. Finiteness and the Nature of Island Constraints. Paper presented at Proceedings of GLOW in Asia IX 2012: The Main Session, Tsu, Japan, September 4–6; Edited by Nobu Goto, Koichi Otaki, Atsushi Sato and Kensuke Takita. pp. 187–97. Available online: https://faculty.human.mie-u.ac.jp/~{}glow_mie/IX_Proceedings_Oral/12Michel&Goodall.pdf (accessed on 30 March 2022).
Müller, Christiane. 2017. Extraction from Adjunct Islands in Swedish. *Norsk Lingvistisk Tidsskrift* 35: 67–85. Available online: http://ojs.novus.no/index.php/NLT/article/view/1399 (accessed on 30 March 2022).
Müller, Christiane. 2019. *Permeable Islands. A Contrastive Study of Swedish and English Adjunct Clause Extractions*. Lund: Lund University.
Müller, Christiane, and Clara Ulrich Eggers. 2022. Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English. *Languages* 7: 125. [CrossRef]
Narita, Hiroki. 2011. Phasing in Full Interpretation. Ph.D. dissertation, Harvard University, Cambridge, MA, USA. Available online: https://ling.auf.net/lingbuzz/001304/current.pdf?_s=QGGWd6bwcPrGEIcu (accessed on 30 March 2022).
Nyvad, Anne Mette, Ken Ramshøj Christensen, and Sten Vikner. 2017. CP-Recursion in Danish: A *c*P/CP-Analysis. *The Linguistic Review* 34: 449–77. [CrossRef]
Palmer, Frank R. 2001. *Mood and Modality*, 2nd ed. Cambridge: Cambridge University Press. [CrossRef]
Pesetsky, David. 1987. Wh-in-Situ: Movement and Unselective Binding. In *The Representation of (in)Definiteness*. Edited by Eric Reuland and Alice ter Meulen. Cambridge: MIT Press, pp. 98–129.
Poole, Ethan. 2017. Movement and the Semantic Type of Traces. Ph.D. dissertation, University of Massachusetts, Amherst, MA, USA. [CrossRef]
Pritchett, Bradley L. 1992. *Grammatical Competence and Parsing Performance*. Chicago: University of Chicago Press.
R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 1 January 2022).
Rizzi, Luigi. 1990. *Relativized Minimality*. Linguistic Inquiry Monographs 16. Cambridge: MIT Press.
Rizzi, Luigi. 2013. Locality. *Lingua*, SI: Syntax and cognition: Core ideas and results in syntax. *Lingua* 130: 169–86. [CrossRef]
Ros, Hilke. 2005. Functional Phrase Structure: Integrating Functional Ideas in Generative Adverbial Syntax. Ph.D. dissertation, Ghent University, Ghent, Belgium.
Ross, John Robert. 1967. Constraints on Variables in Syntax. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
Sawada, Miyuki, and Richard Larson. 2004. Presupposition & Root Transforms in Adjunct Clauses. In *Proceedings of NELS 34*. Edited by Matthew Wolf and Keir Moulton. Amherst: GLSA, pp. 517–28.
Sheehan, Michelle. 2013. The Resuscitation of CED. In *NELS 40 Proceedings of the 40th Annual Meeting of the North East Linguistic Society*. Edited by Seda Kan, Claire Moore-Cantwell and Robert Staubs. Amherst: GLSA, pp. 135–50.
Snyder, William. 2000. An Experimental Investigation of Syntactic Satiation Effects. *Linguistic Inquiry* 3: 575–82. [CrossRef]

Snyder, William. 2022. On the Nature of Syntactic Satiation. *Languages* 7: 38. [CrossRef]
Sprouse, Jon. 2007. Continuous Acceptability, Categorical Grammaticality, and Experimental Syntax. *Biolinguistics* 1: 123–34. [CrossRef]
Sprouse, Jon. 2009. Revisiting Satiation: Evidence for an Equalization Response Strategy. *Linguistic Inquiry* 40: 329–41. [CrossRef]
Sprouse, Jon, Carson T. Schütze, and Diogo Almeida. 2011. Assessing the Reliability of Journal Data in Syntax: Linguistic Inquiry 2001–2010. *Lingua* 134: 219–48. [CrossRef]
Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental Syntax and the Variation of Island Effects in English and Italian. *Natural Language & Linguistic Theory* 34: 307–44. [CrossRef]
Sprouse, Jon, Matt Wagers, and Colin Phillips. 2012. A Test of the Relation between Working-Memory Capacity and Syntactic Island Effects. *Language* 88: 82–123. [CrossRef]
Starke, Michal. 2001. Move Dissolves into Merge: A Theory of Locality. Ph.D. dissertation, University of Geneva, Geneva, Switzerland.
Szabolcsi, Anna. 2006. Strong vs. Weak Islands. In *The Blackwell Companion to Syntax*. Malden: Blackwell Publishing, pp. 479–531. [CrossRef]
Szabolcsi, Anna, and Terje Lohndal. 2017. Strong vs. Weak Islands. In *The Companion to Syntax*. Edited by Martin Everaert and Henk van Riemsdijk. Hoboken: John Wiley & Sons, Inc., pp. 1–51. [CrossRef]
Tanaka, Misako. 2015. Scoping out of Adjuncts: Evidence for the Parallelism between QR and WH-Movement. Ph.D. dissertation, University College London, London, UK. Available online: https://discovery.ucl.ac.uk/id/eprint/1463471/ (accessed on 4 April 2022).
Teleman, Ulf. 1967. Bisatser i Talad Svenska [Subordinate Clauses in Spoken Swedish]. In *Svenskt Talspråk: Fem Studier*. Edited by Gösta Holm. Stockholm: Almqvist & Wiksell.
Teleman, Ulf, Staffan Hellberg, Erik Andersson, Lisa Christensen, and Svenska Akademien. 1999. *Svenska Akademiens Grammatik*. Stockholm: Svenska Akademien. Available online: https://svenska.se/grammatik/ (accessed on 25 November 2019).
Tonhauser, Judith, David I. Beaver, and Judith Degen. 2018. How Projective Is Projective Content? Gradience in Projectivity and At-Issueness. *Journal of Semantics* 35: 495–542. [CrossRef]
Truswell, Robert. 2011. *Events, Phrases, and Questions*. Oxford: Oxford University Press. [CrossRef]
Tutunjian, Damon, Fredrik Heinat, Eva Klingvall, and Anna-Lena Wiklund. 2017. Processing Relative Clause Extractions in Swedish. *Frontiers in Psychology* 8: 2118. [CrossRef]
van Urk, Coppe. 2020. Successive Cyclicity and the Syntax of Long-Distance Dependencies. *Annual Review of Linguistics* 6: 111–30. [CrossRef]
Van Valin, Robert D., Jr. 1994. Extraction Restrictions, Competing Theories and the Argument from the Poverty of the Stimulus. In *The Reality of Linguistic Rules*. Edited by Susan D. Lima, Roberta Corrigan and Gregory K. Iverson. Amsterdam: John Benjamins Publishing Company, pp. 243–60. [CrossRef]
Van Valin, Robert D., Jr. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press. [CrossRef]
Van Valin, Robert D., Jr., and Randy J. LaPolla. 1997. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press. [CrossRef]
Venables, William N., and Brian D. Ripley. 2002. *Modern Applied Statistics with S*. Statistics and Computing. New York: Springer. [CrossRef]
Villata, Sandra, Luigi Rizzi, and Julie Franck. 2016. Intervention Effects and Relativized Minimality: New Experimental Evidence from Graded Judgments. *Lingua* 179: 76–96. [CrossRef]
Vincent, Jake W., Ivy Sichel, and Matthew W. Wagers. 2022. Extraction from English RCs and Cross-Linguistic Similarities in the Environments That Facilitate Extraction. *Languages* 7: 117. [CrossRef]
Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer. Available online: https://ggplot2.tidyverse.org (accessed on 13 September 2021).
Wiklund, Anna-Lena, Fredrik Heinat, Eva Klingvall, and Damon Tutunjian. 2017. An Acceptability Study of Long-Distance Extractions in Swedish. In *Language Processing and Disorders*. Edited by Linda Escobar, Vicenç Torrens and Teresa Parodi. Cambridge: Cambridge Scholars Publishing, pp. 103–20.
Witzel, Naoko, Jeffrey Witzel, and Kenneth Forster. 2012. Comparisons of Online Reading Paradigms: Eye Tracking, Moving-Window, and Maze. *Journal of Psycholinguistic Research* 41: 105–28. [CrossRef] [PubMed]

languages

MDPI

Article

Extraction from English RCs and Cross-Linguistic Similarities in the Environments That Facilitate Extraction

Jake W. Vincent *, Ivy Sichel and Matthew W. Wagers

Department of Linguistics, University of California, Santa Cruz, CA 95064, USA; isichel@ucsc.edu (I.S.); mwagers@ucsc.edu (M.W.W.)
* Correspondence: jwvincen@ucsc.edu

Abstract: In the first two decades following Ross's *Constraints on Variables in Syntax*, a picture emerged in which the Mainland Scandinavian (MS) languages appeared to systematically evade some of the locality constraints proposed by Ross, including the relative clause (RC) part of the complex NP constraint. The MS extraction patterns remain a topic of debate, but there is no consensus as to why extraction from RCs should be so degraded in English (compared to MS)—or why it should be so acceptable in MS (compared to English). We present experiment results which indicate that English should be counted among the languages that allow extraction from RCs in at least some environments. Our results suggest a negligible island effect for RCs in predicate nominal environments and a substantially reduced island effect for those in canonical existential environments. In addition, we show that the size of the island effect resulting from extraction from an RC under a transitive verb is substantially reduced when the transitive verb is used to make an indirect existential claim. We present arguments that patterns of RC sub-extraction discovered in Mainland Scandinavian languages are mirrored in English, and we highlight methodological innovations that we believe may be useful for further investigation into this and other topics.

Keywords: islands; relative clauses; island effects; experimental syntax; *wh*-movement; canonical and noncanonical existentials; movement from DP; acceptability judgments

Citation: Vincent, Jake W., Ivy Sichel, and Matthew W. Wagers. 2022. Extraction from English RCs and Cross-Linguistic Similarities in the Environments That Facilitate Extraction. *Languages* 7: 117. https://doi.org/10.3390/languages7020117

Academic Editors: Anne Mette Nyvad and Ken Ramshøj Christensen

Received: 11 January 2022
Accepted: 27 April 2022
Published: 11 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The empirical landscape related to islands and island sensitivity has been gradually shifting since the first discoveries of islands, occasioning new ideas about the general source of island sensitivity, as well as the nature of particular violations. An example of this shift, and the focus of our study, is relative clauses (henceforth RCs), long considered strong islands for extraction[1]. In the first two decades following Ross (1967), a picture emerged in which the Mainland Scandinavian (MS) languages appeared to systematically evade some of the locality constraints proposed by Ross, including the relative clause (RC) part of the complex NP constraint; research into extraction from RC in MS has consistently shown a selective pattern of acceptable extraction, where RCs in some linguistic environments, but not all, facilitate extraction from the RC (Erteschik-Shir 1973; Erteschik-Shir and Lappin 1979; Allwood 1976, 1982; Maling and Zaenen 1982; Taraldsen 1981, 1982). While the MS extraction patterns, and their proper analysis, is a topic of debate (Engdahl 1997; Kush et al. 2013, 2019; Lindahl 2017; Müller 2014, 2015), it remains a mystery why extraction from RCs should be so degraded in other languages (compared to MS). It is also not yet fully clear why it would be more degraded in some linguistic environments, a distribution which has sometimes suggested that the theory of locality be defined at least in part in terms of information structure, or processing limitations and constraints on working memory (Ambridge and Goldberg 2008; Erteschik-Shir 1973; Hofmeister and Sag 2010; Kluender 1992; Kluender and Kutas 1993; Kuno 1987). A pressing set of empirical questions therefore emerges regarding the extent of variation across both of these dimensions: across languages, and within a language, across linguistic environments. To the extent that some languages, such as the MS languages, show a selective

pattern of extraction from RCs, the question we address is whether these environments vary across languages. We focus on English and present experimental evidence for acceptable extraction from English RCs. As we show, the environments in which extraction is most acceptable in English bear a significant resemblance, if not full identity, to environments identified in other languages. Based on this pattern we suggest that RCs in English are weak islands, exactly as in MS and in Hebrew (Nyvad et al. 2017; Lindahl 2014, 2017; Sichel 2018), and that strong island effects arise only in a subset of environments, which we define as presuppositional DPs. Some have argued that RCs which allow sub-extraction are to be characterized in information-structural terms such as backgroundedness or presupposition (Erteschik-Shir 1973, 1982; Ambridge and Goldberg 2008; Engdahl 1982; Löwenadler 2015). Sichel (2018) argues that the external factors that govern extraction from an RC are no different from those that govern extraction from ordinary DPs: the DP from which extraction takes place must be non-presuppositional.

Presuppositional noun phrases are noun phrases whose denotations have already been introduced into the discourse, sometimes also referred to as *given*. Their referents are presupposed to exist at the point at which the sentence is presented, and the containing sentence asserts that something holds of the referent designated by the presuppositional NP. In contrast, the NP in the pivot of an existential statement, bracketed in (1a), is non-presuppositional, since the sentence is introducing the referent into the discourse, by asserting that it exists. Similarly, the predicative NP following the copula, bracketed in (1b), is also non-presuppostional, since it does not even denote an individual, let alone a presupposed one.

(1) a. There were [posters of the Republican candidate] all over town.

b. Jane Smith was [a good candidate for the job].

There is significant consensus in the literature that extraction from simple NPs, in languages such as English, which allow it, is limited to non-presuppositional NPs (sometimes called *non-specific indefinites* or *non-given* NPs; Bianchi and Chesi 2014; Diesing 1992; Fiengo and Higginbotham 1981). For example, it is easier to extract from a non-presuppositional NP in an existential construction than from a presuppositional NP in an ordinary clause (Moro 1997), in (2). The correlation between presuppositionality and sub-extraction is further observed within the class of direct objects, in the distinction between weak and strong quantifiers (Milsark 1974). NPs with weak quantifiers, such as *many* or *few*, are allowed in the existential construction, whereas NPs with strong quantifiers, such as *each* or *most*, are excluded, in (3). When in direct object position, the former permit sub-extraction much more readily than the latter, in (4).

(2) a. Which candidate$_1$ were [$_{TP}$ there T [$_{vP}$ posters of t$_1$] all over town]?

b. *Which candidate$_1$ were [$_{TP}$ [posters of t$_1$]$_2$ T [$_{vP}$ all over town]]?

(3) a. There were (many/several/few) pictures of Mary on the wall.

b. *There was the/every/each picture of Mary on the wall.

(4) a. Who did you see a picture of?

b. Who did you see many/several/few pictures of?

c. *Who did you see the/each picture of?

d. *Who did you see most pictures of?

In the languages in which it has been attested, extraction from RCs seems to follow a similar, if not identical, pattern. Beyond the known cases in MS, additional acceptable cases of overt extraction from RCs have been attested over the years, in Italian (5c, 7c), Spanish, French, and in Hebrew (5d, 6, 7d). These have been observed in particular environments: when the RC is the pivot of an existential construction, in (5), when the RC is a predicate

nominal, in (6), and when the RC is the direct object of an existential-like transitive construction, dubbed *Evidential Existential* by Rubovitz-Mann (Erteschik-Shir 1973; Rubovitz-Mann 2000, 2012), in (7).[2] And, despite history and appearances, there are reasons to doubt whether English deserves its reputation as a language whose RCs are always strong islands. Instances of extraction in English have surfaced sporadically in the literature, over the years, and they seem to track the same environments, at least impressionistically, as seen in (8a, 8b, 8c) (Chung and McCloskey 1983; Kuno 1976; McCawley 1981).

(5) a. Det er der mange der kan lide.
that are there many who like
'There are many who like *that*'. (Danish; Erteschik-Shir and Lappin 1979, p. 55)

b. Det språket finns det många som talar.
that language exist it many that speak
'There are many who speak that language'. (Swedish; Engdahl 1997, p. 13)

c. Ida, di cui non c'è nessuno che sia mai stato innamorato …
'Ida, whom there is nobody that was ever in love with, …'
(Italian; Cinque 2010, p. 83)

d. Al lexem Saxor, yeS rak gvina axat Se-keday limraox.
on bread black be only cheese one that-worthy to.spread
'On black bread, there is only one cheese that's worth spreading'.
(Hebrew; Sichel 2018, p. 357)

(6) Al ha-haxlata ha-zot, yair lapid haya ha-axaron Se-yada
about the-decision this, Yair Lapid was the-last that-knew
'About this decision, Yair Lapid was the last to know'. (Hebrew)

(7) a. Det kender jeg mange der kan lide.
that know I many who like
'That I know many who like'. (Danish; Erteschik-Shir and Lappin 1979, p. 55)

b. [En sådan frisyr] har jag aldrig sett någon som ser snygg ut i.
that such hairstyle have I never seen anyone who looks good in
'That kind of hairstyle, I have never seen anyone who looks good in'.
(Swedish; Engdahl 1997, p. 24)

c. Giorgio, al quale non conosco nessune che sarebbe disposto ad affidare i propri risparmi …
'Giorgio, whom I don't know anybody that would be ready to entrust with their savings …' (Italian; Cinque 2010, p. 83)

d. Me-ha-sifria ha-zot, od lo macati sefer exad Se-keday le-haS'il
from-the-library this yet not found.I book one that-worth to.borrow
'From this library, I haven't yet found a single book that's worth borrowing'.
(Hebrew; Sichel 2018, p. 358)

(8) a. This is the child who there is nobody who is willing to accept.
(English; Kuno 1976, p. 423)

b. This is the one that Bob Wall was the only person who hadn't read.
(English; McCawley 1981, p. 108)

c. That's one trick that I've known a lot of people who've been taken in by.

(English; Chung and McCloskey 1983, p. 708)

The goal of this study is to confirm this impression experimentally, by systematically manipulating these three contexts: pivot of an existential, predicate nominal, and object of an existential-like construction. To the extent that we find that the pattern of extraction in English replicates the pattern in Scandinavian, Romance, and Hebrew, we will have provided new evidence for the weak island status of English RCs; and we will also have provided new evidence for the cross-linguistically uniform relationship between the presuppositional status of the containing NP and strong islandhood. In a recent study of acceptable extraction from English RCs, Christensen and Nyvad (2022) examine whether English speakers show some of the same selective patterns of RC extraction that speakers of Scandinavian languages do, including sensitivity to lexical frequency, improvement over trials, and a preference for topicalization over *wh*-extraction. They reason that selectivity with respect to extraction suggests that RCs are weak islands, as has been argued for MS, since weak islands allow extraction, selectively. Since they do not find the same effects in English, they conclude that in English, RCs are strong islands, blocking all extraction categorically. By the same token, the finding that English sub-extraction tracks the presuppositionality of the NP as in other languages will suggest (a) that English RCs are no different, with respect to islandhood, from Scandinavian, Romance, and Hebrew, and (b) that English RCs are weak islands. Furthermore, the effect of presuppositional NPs on sub-extraction, observed with simple NPs as well, can be attributed to a strong island, however analyzed (see Diesing 1992 and Sichel 2018 for an implementation in terms of syntactic position). We return to discuss the theoretical implications of this generalization in the conclusions, where we spell out the consequences for recent ideas about acceptable extraction from NP islands (Abeillé et al. 2020; Kush et al. 2019). This paper is organized as follows: Section 2 introduces the study of islands in experimental syntax; Section 3 describes the experiments; Section 4 is the discussion of our results and their potential implications; and Section 5 concludes.

2. Experimental Syntax of Islands

Islands are typically complex syntactic environments, embedded in complex syntactic environments, or both. This makes it a challenge to interpret the acceptability of a sentence containing an extraction from a purported island, because any judgment of acceptability is affected not only by how island-specific constraints affect grammaticality but also by any general contributors to the complexity of the sentence that affect parsability. In this study, we follow the design strategy first devised by Sprouse (2007), and elaborated in Sprouse et al. (2012), which uses a factorial experimental design to decompose the acceptability of an island extraction first into any plausible contributors to degraded acceptability that are not specific to island extraction, and then into how much is "left over" for an island constraint to explain.

We illustrate this approach with a *whether*-island in English, as in (9). Imagine a controlled acceptability judgment experiment in which participants assigned ratings to sentences along a 1–6 Likert-type scale, where 1 is least acceptable and 6 is most acceptable. Suppose that sentences such as (9) received, on average, a rating of 2.

(9) What do you wonder whether John bought? ⟨2⟩

This is a low rating, which could be attributed to a grammatical constraint that is violated by extracting the *what* phrase across *whether*. However, other characteristics of (9) could lead to degraded acceptability, including the mere presence of a *whether*-clause complement and the fact that a long filler-gap dependency spans two clauses. Neither of these characteristics alone violates a grammatical constraint, but each independently increases the syntactic or semantic complexity of the sentence and each thus plausibly decreases its overall acceptability. If instead of measuring the acceptability of only island-

containing sentences (9), we also measure the acceptability of related sentences, then we can estimate and account for these independent contributions to acceptability.

The set of sentences in (10) realizes a 2 × 2 factorial design that relates sentences along two relevant dimensions: Dependency Length (*Short*, *Long*) and Structure (*Island*, *Non-Island*). Square brackets mark the potential island domain, and an underscore marks the gap site; hypothetical average ratings are given in angle brackets in the right margin. Notice that (10d), in the *Long*, *whether-clause* condition, is just (9).

(10) a. *Short*, *that*-clause
Who ___ thinks that John bought a car? ⟨6⟩

b. *Long*, *that*-clause
What do you think that John bought ___? ⟨4⟩

c. *Short*, *whether*-clause
Who ___ wonders [whether John bought a car]? ⟨5⟩

d. *Long*, *whether*-clause
What do you wonder [whether John bought ___]? ⟨2⟩

Ratings from sentences that follow the design in (10) can be used to isolate effects that are specific to extraction from an island. The ratings difference (10a)–(10b) shows that there is a cost of processing a long-distance dependency on acceptability: $6 - 4 = 2$. The ratings difference (10a)–(10c) gives the acceptability cost of embedding via *wonder whether* vs. *think that*: $6 - 5 = 1$. Adding these two costs together, $2 + 1 = 3$, lets us predict how degraded the acceptability of (10d) should be relative to (10a), *if it were only due to the independent costs of Dependency Length* and *Structure*. Under a hypothesis of independent costs, then we should expect (10d) to receive an average rating of 3, i.e., 6 – 3. But the average rating of (10d) indicates that we have an unexplained deficit: it is one point lower than predicted. This 1-pt "deficit" provides an estimate of the island effect.

Sprouse et al. (2012) used the term 'DD score', as in difference of differences, to refer to how much more was needed to explain the low acceptability of an island-containing sentence. In designs such as (10) that manipulate a Length factor with some Structure factor that has *Simple* and *Complex* levels, such as *Non-Island* and *Island* in the example above, the DD score is always defined as the differences between D1 and D2, where D1 represents *Long Simple–Long Complex*, and D2 represents *Short Simple–Short Complex*. This yields a measure that is easy to interpret: if there is an island effect, DD will be positive. In the example above, DD = 1. The presence of an island effect is thus traced to a *superadditive interaction*, one which can be statistically represented by a regression of the ratings measure on the experimental factors.

The DD score method has been used to test a wide range of island types and languages other than English, including Japanese (Sprouse et al. 2011), Brazilian Portuguese (Almeida 2014), Italian (Sprouse et al. 2016), Hebrew (Keshev and Meltzer-Asscher 2018), Slovenian (Stepanov et al. 2018), Norwegian (Kush et al. 2018, 2019), and Modern Standard Arabic (Tucker et al. 2019). Kush et al. (2018) used a design comparable to (10) to investigate adjunct islands, *whether* islands, subject islands, complex NP islands, and—crucially—RC islands in Norwegian. They found that all island types were characterized by a superadditive interaction, i.e., positive DD score, and that the size of the interaction was comparable across subject, adjunct, complex NP and RC islands; it was smaller for *whether* islands, for which the researchers found considerable inter-speaker variation.

Given the discussion above about the often-observed permeability[3] of RCs in MS, the fact that Kush et al. (2018) found an island effect in Norwegian RCs is highly relevant. However, it does not necessarily contradict the observations above, because they did not systematically manipulate the embedding environment to include positions known to "unlock" the island, such as predicate nominal or existential pivot positions. Instead, the RCs appear to be in the complement position of prepositions and transitive verbs. The

set of sentences in (11) below illustrates one of their RC item sets, which crosses Length (11a/11c vs. 11b/11d) and Structure (11a/11b vs. 11c/11d). Observe that the RC is in the complement position of a preposition, in *snakket med* 'speak with' (11c/11d).[4] Their results provide evidence that RCs, in that environment, are islands for extraction in Norwegian.

(11) a. Hvem ___ trodde at et par kritikere hadde stemt på filmen?
who thought that a few critics had voted for film.DEF
'Who thought that a few critics had voted for the film?'

b. Hva trodde regissøren at et par kritikere hadde stemt på ___?
what thought director.DEF that a few critics had voted for
'What did the director think that a few critics had voted for?'

c. Hvem ___ snakket med et par kritikere [som hadde stemt på filmen]?
who spoke with a few critics that had voted for the
'Who spoke with a few critics that had voted for the film?'

d. Hva snakket regissøren med et par kritikere [som hadde stemt på ___]?
what spoke director.DEF with a few critics that had voted for
'What did the director speak with a few critics that had voted for?'

In a later paper, Kush et al. (2019) also investigated extraction from RCs, but this time, the dependency was not a *wh*-question, as in (11), but topicalization. While they found generally smaller DD scores in this experiment, they nonetheless found a positive and significant island effect for topicalization out of RCs.

The key insight from this research is that we can capitalize on a factorial design to experimentally define an island effect. It is important to make a few provisos, however, about this design. Generally these experiments all cross the factors of *Length* and *Structure*, representing the island effect as their interaction. But note that these factors are merely convenient labels for a general design strategy: what they refer to depends on the experiment in question, as the position and nature of the island under consideration varies. *Length* sometimes, but not always, refers also to position of the gap: this is because the shortest dependency often places a gap in matrix subject position (as in 11a/11c above). *Structure* usually refers to the presence or absence of the island but this is then usually conflated with other lexical items. Thus, in (10), changing from a *that* to a *whether* complement necessitates changing the embedding verb (*think*, versus *wonder*). Likewise, in (11), changing from a CP to a DP complement necessitates changing the embedding verb "think" to "speak with". Therefore, some consideration must be given to how *Length* and *Structure* are realized in any given experiment and—crucially—whether the comparison across levels fairly defines a contrast related to the island constraint in question.

A second proviso concerns statistical interactions. In acceptability judgment experiments, participants are usually making their responses on a rating scale where each number on the scale is essentially meaningless other than it defines an order of "goodness" (or "badness"). On a typical 1–6 Likert-type scale, a participant who judges a sentence '2' is judging it to be more acceptable than a sentence to which they have assigned a '1'. Likewise, a participant who judges a sentence a '4' is judging it to be more acceptable than a '3'. But there is no guarantee that a '4' is as much of an improvement on '3' as a '2' is on '1': in other words, these numbers do not define an interval scale. In some participants and experiments, the judgment '2' might correspond to a much wider range of underlying acceptability than '1', say, but less than '3'. It is possible for a spurious statistical interaction to arise if, for example, lower ratings define a much narrower range of acceptability than higher ratings or vice versa (Dillon and Wagers 2021). This is a familiar problem with statistical interactions, when the measurement scale has an unknown relationship to the underlying cognitive constructs (Loftus 1978; Rotello et al. 2015). Two solutions have been

proposed to this problem: one, *magnitude estimation*, has been largely discarded because its assumptions are not met by acceptability judgments (Sprouse 2011). Another, *z-score transformation by participants*, is widely employed to dampen scale bias effects; but it can still give rise to spurious interactions. However, most researchers are at least implicitly aware of this problem and take care to guard against "ceiling" and "floor" effects, which can give rise to some of the pernicious scale compression problems mentioned above. Dillon and Wagers (2021) advocate for using tools from signal detection theory, such as the receiver-operating characteristic function, which directly takes into account how the scale is used, but in the research we report below, we use cumulative ordinal regression modeling to directly estimate the "width" of each ratings category and thus guard against spurious interactions. In figures and data tables, we report average ratings as if they were numbers, for convenience and comparability to previous research, but the underlying data analysis is ordinal.

3. Experiments

As illustrated above, a simple 2 × 2 Length by Structure experiment can be used to estimate island strength for a single domain. However, by holding the domain constant and manipulating an additional factor—the *environment* in which the domain in question is embedded—we can gain insight into the influence of the surrounding environment on the acceptability of extraction and, hopefully, the permeability of relative clause islands in particular environments.

In this research, we expand the Length by Structure design in this way to estimate the permeability of relative clauses in various environments in English. Following the descriptions of the conditions that facilitate extraction from relative clauses in the Mainland Scandinavian languages and Hebrew, we aimed to examine experimentally whether the facts are parallel at any effect size for English.

3.1. Experiment 1: Syntactic/Semantic Environment

This experiment employs the Length by Structure design to measure the permeability of RCs embedded within two of the three environments discussed in Section 1: the nominal pivot of a canonical existential (exemplified by (5) above) and the nominal complement of a copula (exemplified by (6) above). To allow adequate comparison to non-permeable RCs, we included a third environment: the direct object of a transitive verb. This resulted in a 2 × 2 × 3 experimental design (Length by Structure by Environment).

3.1.1. Participants

Forty-eight participants were recruited on Mechanical Turk, and each participant was paid 5.00 USD for their participation. Participants' data were excluded if their average rating for grammatical fillers was below their average rating for ungrammatical fillers. This resulted in two participants' data being excluded, resulting in a total of forty-six participants' data being included in the analysis.

3.1.2. Materials and Methods

The fully crossed design resulted in 12 conditions per item, a sample of which is provided in Table 1. Thirty-six items were constructed in total. The level of the Environment factor referring to the nominal pivot of an existential environment level is labeled *Existential*; the level referring to the nominal complement of a copula is labeled *Predicate* (as in predicate nominal), and the level referring to the object of a transitive verb is labeled *Transitive object*. In contrast to the experiments that follow it, Experiment 1 tested extraction from a relative clause for *wh*-question formation.

Table 1. Experiment 1 sample item.

	Sentence	Length	Structure	Environment
a	Who ___ thinks that Courtney saw that only one art collector bid on this painting?	Short	Non-island	Transitive object
b	Which painting do you think that Courtney saw that only one art collector bid on ___ ?	Long	Non-island	Transitive object
c	Who ___ thinks that Courtney saw the only art collector who bid on this painting?	Short	Island	Transitive object
d	Which painting do you think that Courtney saw the only art collector who bid on ___ ?	Long	Island	Transitive object
e	Who ___ thinks that Courtney believes that only one art collector bid on this painting?	Short	Non-island	Predicate
f	Which painting do you think that Courtney believes that only one art collector bid on ___ ?	Long	Non-island	Predicate
g	Who ___ thinks that Courtney believes that she is the only art collector who bid on this painting?	Short	Island	Predicate
h	Which painting do you think that Courtney believes that she is the only art collector who bid on ___ ?	Long	Island	Predicate
i	Who ___ thinks that there is only one art collector bidding on this painting?	Short	Non-island	Existential
j	Which painting do you think that there is only one art collector bidding on ___ ?	Long	Non-island	Existential
k	Who ___ thinks that there is only one art collector who bid on this painting?	Short	Island	Existential
l	Which painting do you think that there is only one art collector who bid on ___ ?	Long	Island	Existential

All experiment conditions in every item contained the word *only*. In the *Island* conditions for the *Transitive object* and *Predicate* groups, we used DP-internal *only*, following impressionistic judgments that *only* improves the acceptability of existing sub-extraction examples, such as (8b). In the other conditions, *only* was included to maintain lexical matching to the extent possible. The reason that *only* seems to improve the chances of successful sub-extraction in *the*-DPs may be because it removes part of the presuppositional component that commonly accompanies the use of the definite determiner (see McNally 2008, p. 165).

Seventy-two filler sentences were included in this study, all of which were presented to a participant, regardless of which Latin square list the participant received. Both the mean and the median length for the filler sentences was twelve words. The fillers were a mix of grammatical and ungrammatical declaratives and interrogatives. Including both filler and experimental conditions, each participant viewed and rated 108 sentences, half of which were interrogatives and half of which were declaratives. Because all experiment items contained the word *only*, half of the filler sentences were constructed with the word *only*, which resulted in each participant seeing seventy-two *only* sentences and thirty-six sentences without *only*.

One of the challenges faced by researchers extending the factorial definition of islands to relative clauses is illustrated in all of the non-island conditions in Table 1. In order to accurately gauge the permeability of a relative clause in a particular environment, a non-island equivalent must be identified for each environment that plausibly contains all of the same contributors to degraded acceptability that the island condition does, except for those that are specific to island extraction.[5] For the existential conditions, our plausible non-island replaced the relative clause within the nominal pivot with the present participial phrase commonly found in existentials (Deal 2009). For the predicate nominal conditions, we replaced an embedded copular clause with an embedded non-copular clause. For

the *Transitive object* conditions, we replaced RC-containing DP complements with clausal complements. To maintain lexical similarity within those conditions, the embedded verbs for the *Transitive object* conditions were all capable of taking either a DP complement or a clausal complement (*see*, *hear*, *notice*, *remember*, *recognize*, *find*, *discover*, and *mention*).[6]

3.1.3. Analysis

The reported DD scores were calculated on ratings that were z-scored by participant with filler ratings data.

We fit a mixed-effects ordinal regression model with a cumulative link to the ratings data. A maximal random-effects structure was specified. Rating was set as the dependent variable, and Length, Structure, and Environment type were set as fixed effects.

We assigned the Length and Structure factors sum contrast coding and the Environment factor Helmert contrast coding. The effect of this on the model estimation process was that the *Predicate* and *Existential* levels were compared directly to each other, and their mean was compared directly to the *Transitive object* level. We believed this to be sensible since we had reason to believe that the *Predicate* and *Existential* conditions would pattern more closely with each other than with the *Transitive object* conditions. We refer to the comparison between the *Predicate* and *Existential* levels as the Pred_Exist comparison, and the comparison between the combination of those two levels and the *Transitive object* level as the PredExist_Object comparison.

3.1.4. Predictions

We expected to find main effects at least of Length and Structure. Since the *Island*, *Long* conditions involve extraction from a relative clause, we expect to see an interaction between Length and Structure that collapses across the three Environments. If there is indeed a significant reduction in island effects for the *Predicate* and *Existential* environments (as compared to the *Transitive object* environment), we expect a significant three-way interaction between Length, Structure, and the comparison between the *Transitive object* conditions and the means of the *Predicate* and *Existential* conditions. If the island effects observed in the *Predicate* conditions are substantially different than those observed for the *Existential* conditions, we expect to see an interaction between Length, Structure, and the Predicate–Existential comparison.

3.1.5. Results

The mean raw ratings for Experiment 1 are reported in Table 2 and visualized in Figure 1. The collection of *Transitive object* conditions received the lowest ratings overall, followed by the *Predicate* conditions. We see the expected drop in acceptability ratings in the conditions involving extraction from a relative clause (*Long*, *Island*), but this drop is fairly unremarkable in the *Predicate* conditions, suggesting a reduced island effect at least in that environment.

The DD scores calculated from the z-scored ratings in Table 2 are presented in Figure 2. The DD score for the *Predicate* environment is the lowest, which is expected considering the observation made above about the ratings for this condition. The DD score for the *Existential* environment follows, and the DD score for the *Transitive object* environment is substantially higher than that for either the *Predicate* or *Existential* environments. Readers who wish to scrutinize the DD scores by item that are averaged to produce the DD scores in Figure 2 may refer to Appendix D.

Table 2. Descriptive statistics for Experiment 1 results. Mean is calculated on raw (non-z-scored) ratings.

Length	Structure	Environment	Mean Rating	SD	n	SE
Short	Non-island	Tr. object	3.89	1.51	138	0.128
Long	Non-island	Tr. object	3.67	1.60	138	0.136
Short	Island	Tr. object	3.91	1.52	138	0.130
Long	Island	Tr. object	2.60	1.46	138	0.124
Short	Non-island	Predicate	4.10	1.57	138	0.134
Long	Non-island	Predicate	3.63	1.52	138	0.130
Short	Island	Predicate	4.12	1.54	138	0.131
Long	Island	Predicate	3.30	1.60	138	0.136
Short	Non-island	Existential	5.10	1.14	138	0.097
Long	Non-island	Existential	4.30	1.55	138	0.132
Short	Island	Existential	4.84	1.29	138	0.109
Long	Island	Existential	3.57	1.58	138	0.134

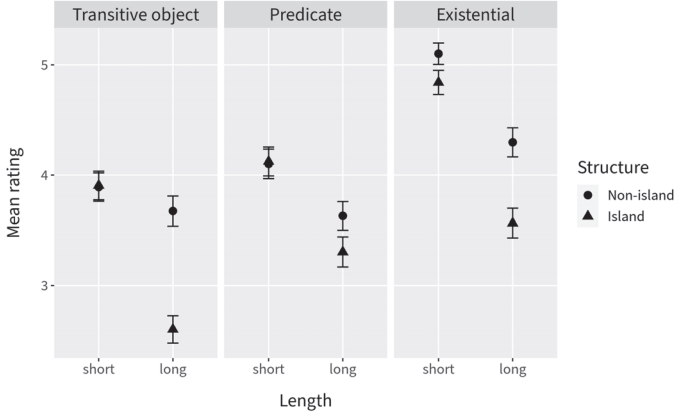


Figure 1. Mean ratings faceted by Environment, arranged in columns by Length. Error bars represent the standard error. Mean is calculated on raw (non-z-scored) ratings.

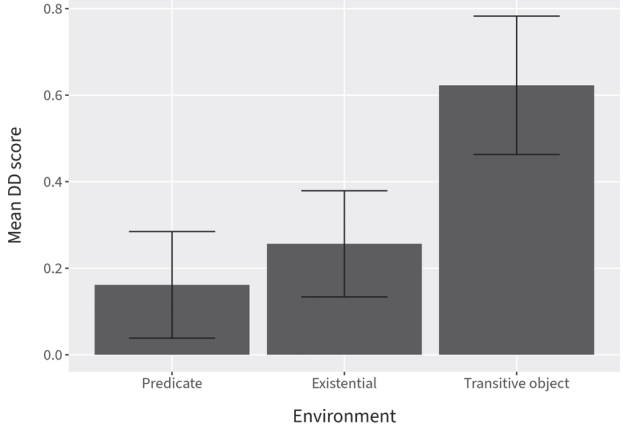


Figure 2. DD scores by Environment (calculated from z-scored ratings). Error bars represent the standard error over DD scores calculated per item. DD scores, left to right: 0.16, 0.26, 0.62. See z-scored ratings by item in Appendix D.1.

In the ordinal regression model (see Appendix E.1 for model output), all environments were significantly different from each other, as revealed by significant main effects of Pred_Exist ($p < 0.001$) and PredExist_Object ($p < 0.001$). Length and Structure also had significant independent effects on ratings (both ps < 0.001). There was a significant island effect overall, as revealed by a significant interaction between Length and Structure ($p < 0.001$).

As hinted at by the relatively low DD scores for the *Predicate* and *Existential* environments (compared to the *Transitive object* environment), there was a significant three-way interaction between Length, Structure, and PredExist_Object ($p < 0.001$). On the other hand, the interaction between Length, Structure, and Pred_Exist was not significant ($p = 0.124$).

3.1.6. Discussion

The results of Experiment 1 suggest that RCs in both the predicate nominal and existential pivot environments are significantly more permeable than RCs in a transitive object environment. The lack of a significant three-way interaction between Length, Structure, and Pred_Exist suggests that the difference between the DD scores for the *Predicate* and *Existential* environments is negligible and that these environments effectively pattern together when it comes to the acceptability of extraction from RCs.

It remains an open question why the DD scores for the two environments that facilitate extraction are above zero. This suggests that there is not a complete amelioration of island effects. However, this finding is reminiscent of Kush et al. (2019), which found residual island effects for most of the island types they examined in Norwegian (despite informal reports of non-islandhood).

3.2. Experiment 2: Existential-like Transitive Verbs (with Supporting Context)

Although Experiment 1 demonstrates a clear reduction in island effect size for predicate nominal and existential environments, the results do not tell us *why* those environments facilitate extraction from RCs in English. The effect could in principle be unique to precisely those two environments, but it could also be due to properties those two environments have in common—properties that other environments might also have.

One property that these two environments have in common is that the DP that contains the RC is non-presupposed. In existential environments, the existence of the referent of the DP pivot is not presupposed because its existence is being asserted. Similarly, in predicate nominal environments, the existence of the referent of the DP predicate is not presupposed; it is asserted in positive predications and denied in negative predications. To say whether it could be the non-presuppositional nature of the DP in these environments that supports extraction or whether something else about these environments is responsible for the effect, one might consider whether transitive verbs that can be *used* in an existential way to introduce a referent—and therefore do not presuppose their direct object—can be counted among the environments that facilitate extraction in English. Rubovitz-Mann (2000) terms such verbs, when co-occurring with a first-person subject, "Evidential Existential" because, as noted in the introduction, the speaker can use them to assert (or deny) the existence of the entity denoted by the direct object by indicating the source of evidence for the existential claim (e.g., in the right context, *I talked to someone who can fix your leak* ≈ *There is (indeed) someone who can fix your leak; I know because I talked to them*). Of course, existential-like transitive verbs are also known to facilitate extraction in the Mainland Scandinavian languages (Engdahl 1997; Erteschik-Shir 1973) and Hebrew (Rubovitz-Mann 2000; Sichel 2018), so examining extraction from RCs in these environments in English is required for a complete picture of the parallels between extraction in English and extraction in the Mainland Scandinavian languages, Hebrew, and the Romance languages.

Because evidential existentiality is basically a pragmatic notion rather than a syntactic notion, a means to measure the compatibility of a transitive verb with an existential use is required—both to determine which transitive verbs should be counted as evidential existentials in an experiment and to determine which should be counted as being incom-

patible with such a use. In a norming rating study, we gauged the compatibility of fourteen transitive verbs with an evidential existential use by presenting a context-setting existential question alongside an affirmative answer that contains one of the following fourteen transitive verbs with a first-person subject: *slap*, *imitate*, *describe*, *criticize*, *advise*, *praise*, *call*, *date*, *run into*, *meet*, *find*, *know*, *hear of*, and *talk to*. A sample dialogue is provided in Figure 3.

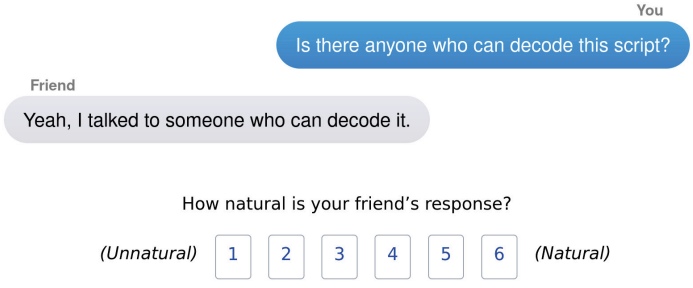


Figure 3. Screenshot of in-experiment dialogue from evidential existentiality norming study.

To ascertain where the felicitousness of the transitive verbs lay with respect to a canonical existential response, we included *there is* as a baseline condition. Our findings are presented in (12), which orders the transitive verbs (and canonical existential) from most to least felicitous under an attempted evidential existential use. The details of the norming study are presented in Appendix B.

(12) *talk to* > *hear of* > *there is* > *know* > *find* > *meet* > *run into* > *date* > *call* > *praise* > *advise* > *criticize* > *describe* > *imitate* > *slap*

The verbs selected for the current experiment were the four transitive verbs rated as most felicitous under an evidential existential use and the four verbs rated as least felicitous. These eight verbs and their z-scored ratings from the norming study are visualized in Figure 4. For comparison, the felicitousness rating of the canonical existential is included in the figure as a horizontal green line.

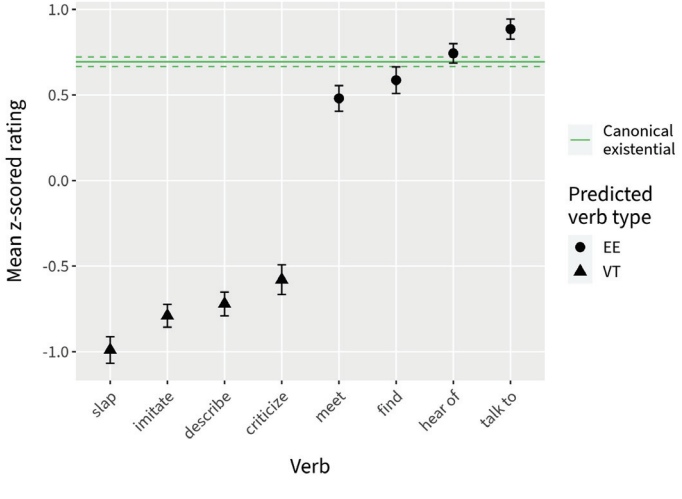


Figure 4. Mean z-scored ratings representing the felicitousness of making an existential claim with eight different matrix verbs (x-axis) in response to an existential question. Error bars (and dashed horizontal lines) represent the standard errors.

The present study utilizes the factorial definition of islands to measure the size of the island effect caused by extraction from RCs under evidential existential transitive verbs (henceforth, EE) and "ordinary" transitive verbs (henceforth, VT). Because the evidential existential use requires a supporting context—one in which the existence of some individual or class of individuals is under discussion and in which the speaker's evidential basis for making an existential claim is necessary or relevant (Rubovitz-Mann 2012, chap. 3)—our goal in developing the materials and methods for Experiment 2 was to supply a context without suggesting to our participants that each declarative sentence was to be judged according to how well it fit in the supplied context. That is, we wanted to ensure that the task was still nominally about judging the acceptability of individual sentences but allow the suggested context to "prime" an evidential existential use of the declarative sentence.

The method we devised was to present a context-setting interrogative as if it were an independent trial to be judged by the participant in the same way as all other trials in the experiment. Normally, trials are randomized or pseudo-randomized in an experiment, so to ensure that the interrogative was capable of suggesting a context for the relevant declarative sentence, we hard-coded the ordering of question trials and their relevant answer trials to ensure that the question had the best chance of implicitly reminding the participant of a possible evidential existential interpretation of the following declarative. Additional details are provided in Section 3.2.2.

3.2.1. Participants

Forty-four participants were recruited for Experiment 2 on Prolific Pro (2022). Participants received 7.13 USD (12.04 USD/h on average) in compensation for their participation. The following exclusion criteria were pre-defined:[7]

(13) Participants will be excluded if at least one of the two following conditions are met:

a. At least 25% of the participant's response times were shorter than one second.

b. The participant's mean ratings for unacceptable and acceptable fillers are either inverted or are too close. *Too close* is defined on normalized (z-scored) ratings as a difference between the average of unacceptable fillers and the average of acceptable fillers that is more than two standard deviations below the mean difference (across participants).

Two participants met the second criterion, and their results were excluded from the analysis, resulting in a total of forty-two participants' data being used. Of the participants whose data were included, their ages ranged from 19 to 71 years. The mean age was 36.1; the median age was 31. Participants were pre-screened so that they could not participate if they had previously participated in experiments run on Prolific for this research. They were required to be born in and currently reside in the United States and were required to have English as their first language or as one of two first languages. They were required to not have any language-related disorders and to have received at least a high school diploma.

3.2.2. Materials and Methods

The materials for Experiment 2 were constructed according to a reduced factorial design. As in Experiment 1, three factors were crossed: Length (*Short*; *Long*), Structure (*Non-island*; *Island*), and now, Verb type (*EE*; *VT*). In this and the following experiment, the sentences presented for judgment were not *wh*-questions (in contrast to those for Experiment 1) but declaratives involving relativization. This move was made so that we could utilize a context-setting interrogative, which would provide the context for the critical conditions. A full factorial design would have resulted in eight conditions per item ($2 \times 2 \times 2$), but because the non-island conditions for the two verb types would have been identical, one duplicate set of non-island conditions was left out, resulting in six conditions per item. The non-island conditions were given the label *CP* for the verb type factor because

the non-island conditions were all constructed with a CP-complement-taking verb (one of *believe*, *claim*, *imagine*, *suggest*, *suspect*, or *think*).

Each condition consisted of a pair of sentences: a context-setting question and a relevant answer to that question. The questions were existential in nature, each one asking whether any individual who meets the conditions described in a restrictive relative clause exists. The answers to these questions were all declarative statements that could be taken as indirect existential assertions in response to the question. A sample item for Experiment 2 is provided in Table 3.

Table 3. Experiment 2 sample item.

	Sentence	Length	Structure	Verb Type
	Preceding question: Is there anyone who could prove this claim?			
a	This is the woman that ___ just realized that an expert could prove this claim.	Short	Non-island	CP
b	This is a claim that I just realized that an expert could prove ___.	Long	Non-island	CP
c	This is the woman that ___ just found an expert who could prove this claim.	Short	Island	EE
d	This is a claim that I just found an expert who could prove ___.	Long	Island	EE
e	This is the woman that ___ just slapped an expert who could prove this claim.	Short	Island	VT
f	This is a claim that I just slapped an expert who could prove ___.	Long	Island	VT

Thirty-six items were constructed in total, twelve of which were reserved for an initial practice period that we henceforth refer to as a "burn-in" practice period.[8] Trials from the burn-in practice period ("burn-in trials") were not analyzed. The purpose of including burn-in trials is to ensure that the data included in the analysis were acquired after participants had acclimatized to the ratings scale and the variety of sentences they would be judging. As shown in Figure 4, the four verbs used for the *VT* conditions were *slap*, *imitate*, *describe*, and *criticize*; and the four verbs used for the *EE* conditions were *meet*, *find*, *hear of*, and *talk to*. These were distributed equally across the items (each verb was used in six different items).[9] Ratings data were collected for one item whose EE conditions were found to have a typo.[10] Because the typo was discovered after data collection, the ratings for this item were excluded from all analysis. This resulted in considering one less data point per participant than intended.

To prevent participants from judging the acceptability of the sentences *qua* answers to the questions, the task instructions asked participants to focus on the acceptability of each individual sentence. However, Q–A pairs were treated as a unit for Experiment 2, by which we mean that when a question was presented for a participant to rate, the relevant answer was always next in line to rate. As a result, any effect of context on the acceptability of extraction from a relative clause is expected to be implicit, rather than the simultaneous presentation of question and answer as a dialogue. In addition to this structure imposed on the order of question trials and relevant answer trials, we coded a 500 ms separator between all trials except adjacent question trials and a relevant answer trial. These had no separator, so upon selecting an acceptability rating for the question, the participant would immediately be presented with the relevant answer (see the visualization of the placement of the 500 ms separator in Figure 5).

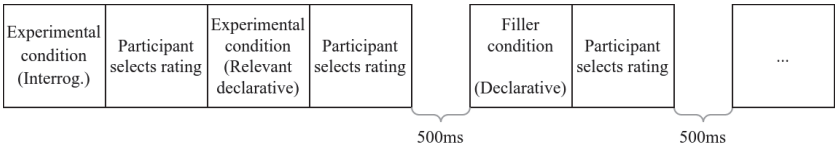


Figure 5. Trial order structure in Experiment 2, highlighting placement of 500 ms separators.

Three sets of fillers were constructed with the goal of ensuring a relatively even balance of grammatical and ungrammatical interrogatives and declaratives and a selection of paired (i.e., adjacent) interrogatives and declaratives, isolated interrogatives, and isolated declaratives. A total of 126 filler items were constructed in total, forty-two of which were reserved for the "burn-in" practice period. Approximately 26% of trials overall were grammatical interrogatives; 18% were ungrammatical interrogatives; 29% were grammatical declaratives; and 27% were ungrammatical declaratives. Out of all trials, approximately 34% were interrogatives adjacent to a relevant declarative, 34% were declaratives following a relevant interrogative, 10% were isolated interrogatives, and 22% were isolated declaratives.

As noted above, burn-in items were created for both experimental and filler items. A period lasting for about the first third of the experiment (about 100 trials, twelve of which were from the experimental items) was dedicated to the burn-in items. In the interest of transparency, descriptive statistics from the experimental burn-in trials are provided in Appendix C.

For instructions on how to access a working demonstration copy of Experiment 2, please see Appendix A.

3.2.3. Analysis

To derive the DD scores presented below, we calculated z-scores by participant using the ratings data for the main experimental and filler conditions following the burn-in period.

We fit a mixed-effects ordinal regression model with a cumulative link to the ratings data. A maximal random-effects structure was specified. Rating was set as the dependent variable, and Length and Verb type were set as fixed effects. Again, Structure was not included in the analysis because the reduced structure of the experiment design, combined with the contrast coding given to the Verb type factor, resulted in Structure not providing any independent information.

We assigned the Length factor sum contrast coding and the Verb type factor treatment contrast coding. This effectively treats the CP-complement level as the baseline condition for the other two verb types. For this factor, this results in the *EE* and *VT* conditions not being compared directly to each other, but to the other condition's difference with the *CP* level.

3.2.4. Predictions

We anticipated main effects of Length (*Short* > *Long*), Structure (*Non-island* > *Island*), and Verb type (*EE* > *VT*). Main effects for Length and Structure are expected because of the greater processing demands involved in processing longer-distance (vs. shorter-distance) dependencies and in processing embedded clauses requiring filler-gap resolution (vs. those that do not). We expect a main effect of Verb type because the more specific meaning of the *VT* conditions was less relevant to the context set by the adjacent question. Due to the treatment contrast coding applied to the Verb type factor, we expect the latter main effect to show up as a significant main effect of *VT* as compared to the *CP* level and an insignificant main effect of *EE* as compared to *CP*.

At the very least, we expect to see a significant interaction between Length and Structure for the *VT* conditions; this would be the standard island effect. If island effects are completely ameliorated for the *EE* conditions, we would not expect to see a significant interaction between Length and Structure for the *EE* conditions. However, considering that there was still a significant interaction between Length and Structure for the *Existential*

conditions in Experiment 1, we may observe a reduction in island effects for the *EE* conditions that does not completely remove the interaction between Length and Structure.

3.2.5. Results

The descriptive statistics are summarized in Table 4, and the mean ratings in Table 4 are visualized in Figure 6. The reader will note that there is a generally reduced acceptability associated with the *VT* conditions, suggesting that the more specific event descriptions of the verbs used in those conditions caused degradation, that these conditions were less acceptable as answers to existential questions, or a mixture of both of these possibilities. Unsurprisingly, the *EE*, *Long* and *VT*, *Long* conditions were the most degraded, falling below long-distance extraction from a complement clause (*CP*).

Table 4. Descriptive statistics for Experiment 2 results. Mean is calculated on raw (non-z-scored) ratings.

Length	Structure	Verb Type	Mean Rating	SD	n	SE
Short	Non-island	CP	4.25	1.35	161	0.106
Long	Non-island	CP	3.81	1.52	161	0.120
Short	Island	EE	4.39	1.33	161	0.105
Long	Island	EE	3.32	1.49	161	0.118
Short	Island	VT	3.86	1.43	161	0.113
Long	Island	VT	2.45	1.32	161	0.104

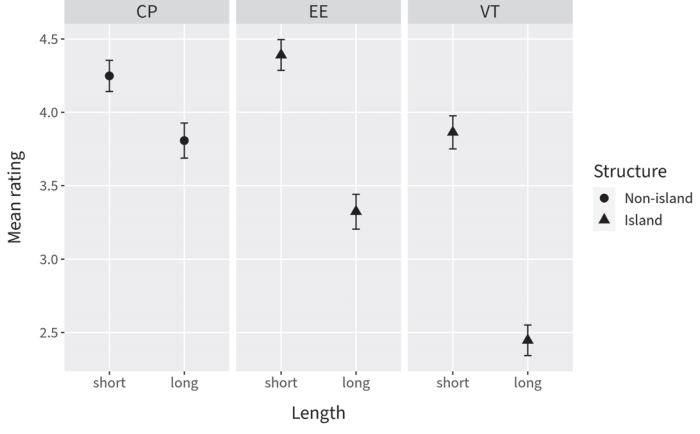


Figure 6. Mean ratings for Experiment 2. Error bars represent the standard error. Mean is calculated on raw (non-z-scored) ratings.

The DD score plot in Figure 7 shows the range of DD scores calculated for each verb used in Experiment 2. The DD scores for the *EE* verbs were lower on average than those for the *VT* verbs, but one verb categorized as *VT* (*criticize*) fell among the lowest DD scores, and one verb categorized as *EE* (*find*) fell among the highest DD scores. Despite these apparent outlier DD scores, we take these DD scores to be a confirmation of our predictions from a descriptive statistical standpoint: RCs within non-presupposed direct objects are more permeable than those within the direct objects of more typical transitive verbs.

In the ordinal regression model (see Appendix E.3 for model output), we observed a main effect of Length ($p = 0.022$). The comparison of the *CP* conditions to the *EE* conditions was just outside of the 0.05 significance threshold ($p = 0.064$), indicating that we cannot reliably conclude that the *EE* conditions were judged any differently from the *CP* conditions overall. On the other hand, the comparison of the *CP* conditions to the *VT* conditions was significant ($p < 0.001$), which is consistent with the impressions given by Figure 6.

Both length interactions were significant ($ps < 0.001$), although the interaction between length and the *CP–EE* comparison received a smaller coefficient estimate, indicating a smaller effect size for that interaction.

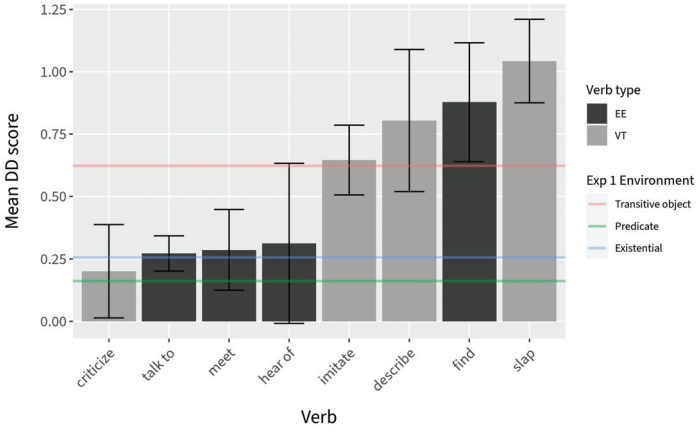


Figure 7. DD scores (calculated from z-scored ratings) by verb for Experiment 2 with DD scores for Experiment 1 environments overlaid as horizontal lines. Error bars represent the standard error over DD scores calculated by item. Summary statistics are based on five to six DD scores calculated for each verb. See z-scored ratings by item in Appendix D.2.

3.2.6. Discussion

The significance of the interactions in the ordinal regression model indicates that even with supporting context, there is still a significant island effect for both verb types. However, both the DD scores and the coefficient estimates for the models indicate a smaller effect size for EE verbs, which suggests that the island effect for that verb type is reduced.

3.3. Experiment 3: Existential-like Transitive Verbs (without Supporting Context)

In order to gauge the impact of the indirectly suggested context on the island effects observed in Experiment 2, we constructed and deployed Experiment 3, which was identical to Experiment 2 except that the context-setting questions were paired with an item whose answers were unrelated and irrelevant. All other aspects of the experiment remained unchanged from Experiment 2.

3.3.1. Participants

Forty-four participants were recruited for Experiment 3 on Prolific. Participants received 7.13 USD (11.26 USD/h on average) in compensation for their participation. The same exclusion criteria were used for Experiment 3 as were used for Experiment 2.

Again, two participants met the second criterion, and their results were excluded from analysis, resulting in a total of forty-two participants' data being included in the analysis. Of the participants whose data were included, their ages ranged from 18 to 64 years. The mean age was 34.7; the median age was 33. Participants were pre-screened so that they could not participate if they had previously participated in experiments run on Prolific for this research. They were required to be born in and currently reside in the United States and were required to have English as their first language or as one of two first languages. They were required to not have any language-related disorders and to have received at least a high school diploma.

3.3.2. Materials and Methods

The materials and methods used for Experiment 3 were identical to those used for Experiment 2, but the question and answer components of each item were scrambled so

that participants would never see a relevant declarative statement that could felicitously be interpreted as an answer to the question in the immediately preceding trial. The task instructions remained the same; participants were instructed to rate the acceptability of each sentence, whether declarative or interrogative, on an individual basis. The 500 ms separator was implemented in exactly the same situations, but due to the scrambling of questions and relevant answers, the lack of a separator was no longer a subliminal cue that an adjacent question and answer might be construed together. A sample item is provided in Table 5; note, in particular, that the associated question is irrelevant to the set of possible answers. Due to the shared materials between Experiment 2 and Experiment 3, data for the same item that had a typo in Experiment 2 were also collected but excluded from all analysis.

Table 5. Experiment 3 sample item.

	Sentence	Length	Structure	Verb Type
	Preceding (irrelevant) question: Is there anyone who could fix this toilet?			
a	This is the person that ___ just realized that a chef could prepare this dish.	Short	Non-island	CP
b	This is a dish that I just realized that a chef could prepare ___.	Long	Non-island	CP
c	This is the person that ___ just found a chef who could prepare this dish.	Short	Island	EE
d	This is a dish that I just found a chef who could prepare ___.	Long	Island	EE
e	This is the person that ___ just slapped a chef who could prepare this dish.	Short	Island	VT
f	This is a dish that I just slapped a chef who could prepare ___.	Long	Island	VT

For instructions on how to access a working demonstration copy of Experiment 3, please see Appendix A.

3.3.3. Analysis

The DD scores presented below were calculated in the same way as for Experiment 2.

We fit a mixed-effects ordinal regression model with a cumulative link to the ratings data from Experiment 3. A maximal random-effects structure was specified. Rating was set as the dependent variable, and Length and Verb type were set as fixed effects. Structure was not included in the analysis because the reduced structure of the experiment design, combined with the contrast coding given to the Verb type factor, resulted in Structure not providing any independent information.

We assigned the Length factor sum contrast coding and the Verb type factor treatment contrast coding. This effectively treats the *CP* level as the baseline condition for the other two verb types. For this factor, this results in the *EE* and *VT* conditions not being compared directly to each other, but to the other condition's difference with the *CP* level.

In order to obtain a more direct comparison of the results from the two experiments, we also pooled the ratings data, introduced an Experiment factor (which we also refer to as Q–A order, with the levels *Fixed*, for Experiment 2, and *Scrambled*, for Experiment 3), and estimated a second mixed-effects ordinal regression model for the pooled data. In the regression formula for this second model, Experiment was coded as an additional factor (see Appendix E.5).

3.3.4. Predictions

We anticipated main effects of Length as well as main effects for both Verb type comparisons. Main effects for Length are expected because of the greater processing demands involved in processing longer-distance (vs. shorter-distance) dependencies and

in processing embedded clauses requiring filler-gap resolution (vs. those that do not). In contrast to our expectations for Experiment 2, we do not expect different main effects of Verb type because the effect of scrambling questions and relevant answers is that no declaratives that follow questions will be felicitous answers. Because one *EE* sentence and one *VT* sentence per item involved extraction from a relative clause and the *CP* conditions did not, we expect main effects of verb type for both the *EE–CP* comparison and the *VT–CP* comparison.

We expect to see a significant interaction between Length and Structure for both the *VT* and *EE* conditions, reflecting an island effect for relative clauses under both Verb types.

3.3.5. Results

The mean ratings data are summarized in Table 6 and visualized in Figure 8. Overall, the results appear quite parallel to the results from Experiment 2, but there was a slight increase in the ratings for both *Non-island* conditions, a decrease in the mean rating for the *EE*, *Long* condition, and an increase in the *VT*, *Short* condition.

Table 6. Descriptive statistics for Experiment 3 results. Mean is calculated on raw (non-z-scored) ratings.

Length	Structure	Verb Type	Mean Rating	SD	n	SE
Short	Non-island	CP	4.58	1.16	161	0.091
Long	Non-island	CP	4.29	1.24	161	0.098
Short	Island	EE	4.35	1.23	161	0.097
Long	Island	EE	3.04	1.31	161	0.103
Short	Island	VT	4.04	1.35	161	0.106
Long	Island	VT	2.45	1.28	161	0.101

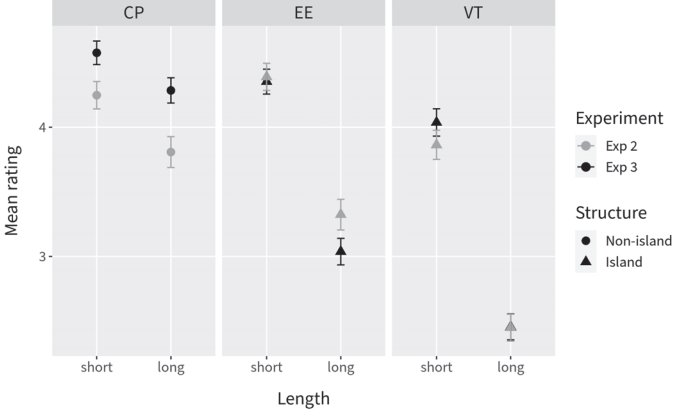


Figure 8. Mean ratings for Experiment 3 (Experiment 2 ratings shown in light gray). Error bars represent the standard error. Mean is calculated on raw (non-z-scored) ratings.

The DD scores calculated by verb for the Experiment 3 data are presented in Figure 9 alongside the DD scores for Experiment 2. Notable differences from the DD scores for Experiment 2 include a disproportionate increase in DD scores for the *EE* verbs except for *talk to*, whose DD score remained basically unchanged. The scores for the *VT* verbs remained fairly constant, but the DD score for *criticize*, which was unexpectedly low in Experiment 2, increased.

In the ordinal regression model we fit to the ratings data, there was a main effect of Length ($p = 0.0079$), and both comparisons (*EE*; *VT*) to the *CP* conditions were significant (ps < 0.001). Additionally, the interactions between Length and the *CP* comparisons were significant (ps < 0.001).

In the analysis of the pooled ratings data from the two experiments, we found a significant main effect of Experiment ($p = 0.008$). See the coefficient estimates for the combined analysis in Appendix E.5. There was a significant interaction between Experiment and the *EE–CP* comparison ($p < 0.001$), and the parallel *VT–CP* comparison interaction was not significant ($p = 0.07$). The interaction between Experiment, Length, and the *EE–CP* comparison was not significant ($p = 0.109$), nor was the interaction between Experiment, Length, and the *VT–CP* comparison ($p = 0.236$).

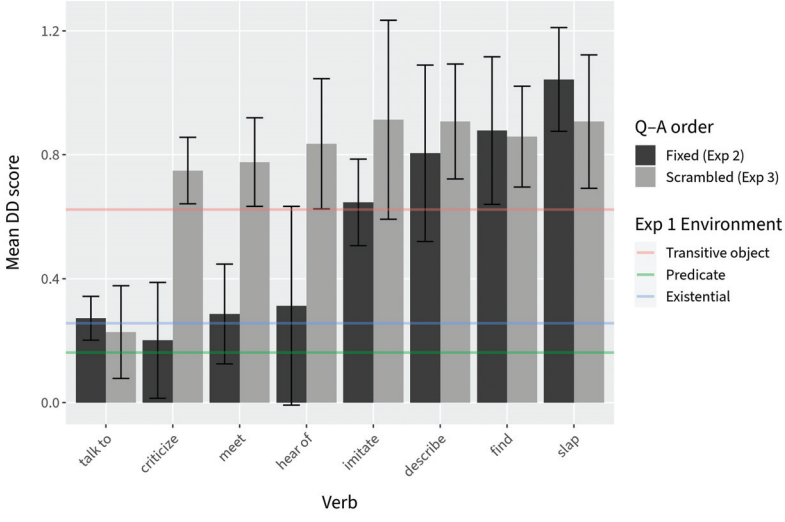


Figure 9. DD scores (calculated from z-scored ratings) by verb and Q–A order (Experiment) with Experiment 1 DD scores overlaid as horizontal lines. Error bars represent the standard error over DD scores calculated by item. Summary statistics are based on five to six DD scores calculated per verb per experiment. See z-scored ratings by item in Appendix D.

3.3.6. Discussion

The *EE–CP* comparison was significant in Experiment 3, in contrast to Experiment 2, which suggests that context has an outsize effect on the acceptability of evidential existential responses compared to typical transitive verbs. In the ordinal regression for the pooled data (in which Experiment was included as a factor), the significant interaction between Experiment and *EE–CP* confirms that this difference across experiments was significant. We take this to be a validation of the notion of an evidential existential use for a transitive verb, as well as the notion that certain verbs more naturally fall into this class than others.

As predicted, we cannot reliably conclude that either the *EE* or *VT* conditions completely lacked an island effect, as indicated by the significant interactions between Length and both *EE/VT–CP* comparisons. The combined ordinal regression model also indicated that the strength of the island effect is not significantly different for either Verb type level across the two experiments, which means we cannot conclude with certainty that context *generally* increased the permeability of RCs in evidential existential contexts. This is reflected by the closeness of the error bars in the DD score plot presented in Figure 10, which collapses DD scores by Verb type. Although the slight non-overlap of the error bars in the *EE* half of the plot, along with the slight overlap of the error bars in the *VT* half of the plot, gives the impression of a disproportionate effect of context on RC permeability for the *EE* conditions (as predicted), the data do not allow us to conclude with confidence that this is the case.

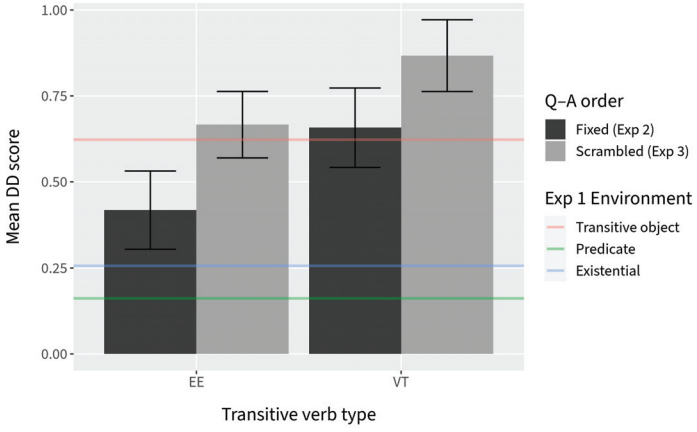


Figure 10. Average DD scores (calculated from z-scored ratings) by transitive verb type and Q–A order (Experiment) with Experiment 1 DD scores overlaid as horizontal lines. Error bars represent the standard error over DD scores. See z-scored ratings by item in Appendix D.

4. General Discussion

The inferential statistics for Experiments 2 and 3 indicate a persistent interaction between Length and Environment, regardless of Verb type. Taking these results seriously, we cannot conclude that there was a complete absence of island effects in either experiment. This conclusion is confirmed by the ordinal regression model estimated for the combination of the data from the two experiments: the lack of a significant interaction between Length, Environment, and Experiment (for either verb type) indicates that we cannot confidently conclude that there was a significant difference in island effect across Experiment 2 and Experiment 3 within each Verb type.

However, examination of the DD scores suggests that the combined effects of Verb type and context are not inconsequential. Although we observed a general increase in the DD scores for both verb types in Experiment 3, the DD scores for the *EE* verb type pull apart slightly more across the two experiments when compared to the *VT* verb type (Figure 10). Further, when the mean DD scores visualized in Figure 10 are broken down according to verb (Figure 9), there are notable trends within each verb type. The only verb in the *EE* group that maintained consistently low DD scores across the two experiments was *talk to*. This is unlikely to be due to chance; the results from the evidential existentiality norming study indicate that out of fourteen transitive verbs tested, *talk to* is the most natural transitive verb with which to make an "evidential existential" claim (for additional discussion, see Vincent 2021). Two of the other four *EE* verbs used in Experiments 2 and 3, *meet* and *hear of*, have a noticeably higher DD score in Experiment 3, when context did not favor an existential use. Similarly, three out of four verbs that were categorized under *VT* (*imitate*, *describe*, and *slap*) maintained consistently high DD scores across the two experiments. This also seems unlikely to be due to chance, as these three verbs were found to be the least natural transitive verbs to use to make an existential claim in a supporting context.

What this suggests to us is that there is a gradient effect on relative clause permeability that is affected by the likelihood of the transitive verb being used existentially. Certain verbs such as *talk to* are so natural in non-canonical existential assertions that a reading in which their complement is non-presupposed is easily accommodated. Verbs such as *imitate*, *describe*, and *slap*, on the other hand, are so unnatural in existential assertions that a non-presupposed reading of their complement is difficult to accommodate—even when context provides the right conditions for an existential assertion. It is also possible that

there is variation across speakers regarding the possibility for a non-canonical existential reading for particular verbs, contributing to the overall less clear picture.

In conjunction with the results from Experiment 1, in which canonical existential and predicate nominal environments result in a substantial decrease in island effects, the picture that emerges is that the same factors appear to modulate RC permeability in English as in the Mainland Scandinavian languages: extraction is facilitated when the RC is within a predicate nominal, an existential pivot, or a direct object of a verb with which it is natural to make an existential assertion (refer to the combined DD score plot in Figure 11). This finding is noteworthy from an empirical standpoint because it contrasts with the general consensus that English islands (apart from *whether*-complements) invariably give rise to severe degradation under extraction.

From a theoretical standpoint, our findings provide some clues as to which analyses of extraction from RC may turn out to be fruitful and which may turn out to be unfruitful. What initially appeared to be a phenomenon specific to the Mainland Scandinavian languages may be a more general pattern than initially thought. If the phenomenon's first discovery in these languages is what initially led to suggestions that island constraints be parameterized to handle cross-linguistic variation, then finding that this phenomenon is observable even in English should take us at least one step away from parameterization. It appears likely that the picture is both more cross-linguistically uniform and also more nuanced, language-internally, than a parameterization approach could satisfactorily handle.

Besides the language-particular effects found in English, another conclusion which emerges from our experiments is that the environments which facilitate extraction seem to be cross-linguistically uniform: extraction is permitted (or more acceptable) from a non-presupposed RC (Erteschik-Shir 1973, 1982; Engdahl 1997; Rubovitz-Mann 2000; Sichel 2018; Vincent 2021). Regardless of the ultimate "island" status of some of these environments, the existence of such a consistent cross-linguistic landscape suggests that there is something to understand about these environments and why they facilitate extraction to the extent that they do. The significance of these particular environments is further highlighted by the fact that sub-extraction from simple, non-relative DPs in English follows the same pattern: possible when DP is a non-presupposed indefinite. Here, too, the English pattern is similar to what is known about other languages (Davies and Dubinsky 2003; Diesing 1992; Fiengo and Higginbotham 1981; Mahajan 1992, among others). This suggests that presuppositional DPs are strong islands, and that English RCs, when non-presuppositional, are weak islands, as in other languages in which sub-extraction is attested. Another empirical benefit of our study is that it provides a clear blueprint for future studies in other languages: measurement of sub-extraction facilitation effects depends on knowing where to look for them. Rather than comparing, for example, extraction from RC in subject position vs. extraction from RC in object position, or extraction from indefinite RCs vs. definite RCs, it seems to us that, to the extent that it is at all possible in a language, sub-extraction from an RC is most likely to be found in the sort of non-presuppositional contexts we have focused on.[11] Further investigation of these environments in other languages is needed for a clearer understanding of the cross-linguistic landscape of RC island-hood and its relationship to general DP island-hood.

On the theoretical side, a more nuanced conception of the environments which facilitate sub-extraction is key for the analysis of these cases and for our understanding of the nature of island violations more generally. First, the claim in Sichel (2018) that the external environments which facilitate RC sub-extraction are no different from those which support sub-extraction from simple DPs is further supported by the English pattern. If this is so, and to the extent that sub-extraction from simple DPs can ultimately be analyzed in terms of the syntactic position (derived, non-derived) of presupposed and non-presupposed DPs (Bianchi and Chesi 2014; Diesing 1992), there is no a priori reason to suspect that sub-extraction from RCs is any different: an RC from which extraction is acceptable is in a non-derived position, consistent with contemporary theories of DP-islandhood, which

allow sub-extraction from a simple DP when that DP is in a non-derived position (Rizzi 2004; Stepanov 2001; Takahashi 1994; Uriagereka 1999; Gallego and Uriagereka 2006, 2007; Chomsky 2008; among others).

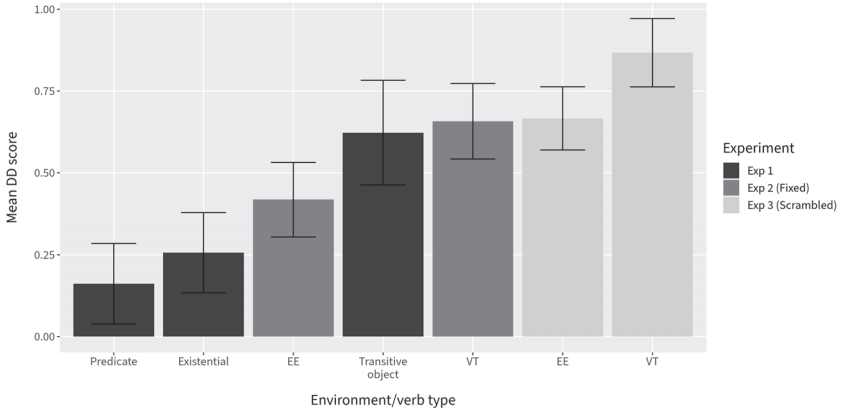


Figure 11. DD scores (calculated from z-scored ratings) across the three experiments reported in this work. Error bars represent the standard error. See z-scored ratings by item in Appendix D.

Second, the empirical cut which emerges from English, along with other languages which permit RC sub-extraction to some degree, can be used to further test predictions raised by other theories of acceptable extraction from islands. In a recent paper on extraction from subject islands, Abeillé et al. (2020) focus on the nature of the extracted constituent and argue for an information-structure based constraint on sub-extraction from subjects, according to which extraction is subject to a focus-background conflict constraint (FBC), a gradient constraint disallowing a focused element to be part of a backgrounded constituent. They compared A-bar extraction for *wh*-questioning with A-bar extraction for relativization, across subjects and objects. They found that extraction from a subject is degraded compared to extraction from an object when extraction is part of question formation—but not when it is part of relativization. The effect is attributed to a clash between the focus potential of the *wh*-phrase and the givenness of subjects, generally. While we basically agree with the characterization of the extraction domain which hinders sub-extraction in terms of information structure, and with the specific characterization in terms of pre-suppositionality (or *givenness*, in the terms of Abeillé et al. 2020), we believe that our more nuanced approach to the distribution of these environments is helpful for further testing of their predictions. While Abeillé et al. (2020) have characterized the overall difference between subjects and objects in terms of givenness, we follow contemporary findings in syntax and semantics which acknowledge that presuppositionality has an effect on sub-extraction both *within* the domain of subjects, as well as *within* the domain of objects: presupposed subjects, as well as presupposed objects, block sub-extraction, whereas non-presupposed objects, as well as non-presupposed subjects, are more porous for sub-extraction. We also think that it is premature to attribute this sensitivity to a clash between the information-structural properties of the extraction domain and the information-structural properties of the extracted constituent. If the source of the problem were indeed such a clash, the expectation is that the characterization of the extraction domain should vary across extraction types—and should reverse when the extracted constituent is information-structurally characterized as *given*, or presupposed. In particular, the types of A-bar movement which apply to given, presuppositional constituents, such as scrambling and topicalization, should actually be more acceptable when the extraction domain is a presupposed (or given) DP than when it is non-presupposed. Our own study used both *wh*-movement in question formation (Experiment 1) and relativization (Experiments 2 and 3) and made no attempt to manipulate

them systematically. Kush et al. (2019) found a lower penalty for topicalization out of RCs than for *wh*-questioning out of RCs but made no attempt to systematically manipulate environments which 'unlock' islands. Sichel (2018) found that topicalization from an RC follows the same presuppositional pattern as in the present study, an indication that the extraction domain does not vary with the information-structure characterization of the extracted constituent. That study, however, is not experimental and did not include the careful quantitative controls that experimental studies, such as the former studies, do. We hope that future work will test these comparative predictions by combining careful quantitative controls and nuanced manipulation of the blocking and facilitating environments.

Although less central to the main focus of this paper, we hope to impress two main methodological points upon our readers. First, we believe that our experiments can be viewed as a trial of the Length by Structure experiment design and an example of how it can be extended to measure not only the permeability of individual island domains but the influence of additional factors (such as environment and context) on the permeability of island domains. Second, we believe that our effort to *suggest* a context (in Experiment 2) without changing the nature of the acceptability judgment task was successful, considering the distinctions we observed in the results for experiments that were identical except for the relevance of Q–A pairs. Future research in this and other areas may find this technique useful when context is relevant or is part of an experiment manipulation but when it is undesirable to directly ask participants to consider an item with respect to a context.

5. Conclusions

Our results indicate that English should be counted among the languages that allow extraction from RCs in at least some environments. The results from Experiment 1 suggest a negligible island effect for RCs in predicate nominal environments and a substantially reduced island effect for those in canonical existential environments. The interactions between the Environment comparisons and Length were significant in both Experiments 2 and 3, indicating that the data collapsed across verbs still bear the signature of a significant island effect. However, the DD scores calculated by verb reveal a somewhat more complex story: the scores for three out of four of the verbs we categorized as *EE* verbs (*talk to*, *meet*, and *hear of*) are on a par with the DD score for canonical existentials in Experiment 1 when participants are "primed" by an adjacent context-setting question.

In addition to the above findings, an important takeaway is that cross-linguistically, the factors that enhance a relative clause's permeability appear to be stable, even if the size of their effects on acceptability ratings vary somewhat. It is a clear pattern that environments and contexts that support existential, non-presupposed interpretations of the DP containing the RC 'unlock' the RC to some extent, whether the environment is a direct assertion (or denial) of existence, a nominal predication, or an indirect assertion (or denial) of existence using an evidential existential verb in a supporting context.

Lastly, we highlighted the methodological innovations that we believe may be useful for further investigation into this and other topics. These include expansion of the Length by Structure design to compare extraction environments as closely as possible as well as the use of trial adjacency to suggest interpretation and evaluation of a condition in the context of another condition without disturbing the overall task.

Author Contributions: Conceptualization, J.W.V., I.S. and M.W.W.; Data curation, J.W.V.; Formal analysis, J.W.V., I.S. and M.W.W.; Methodology, J.W.V., I.S. and M.W.W.; Visualization, J.W.V.; Writing—original draft, J.W.V., I.S. and M.W.W.; Writing—review & editing, J.W.V., I.S. and M.W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study is exempt from IRB review as determined by the Institutional Review Board of the University of California, Santa Cruz (protocol code HS0801386, granted 09/04/2009).

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: The data presented in this study are openly available at the Open Science Framework: https://osf.io/tz7af (accessed on 5 May 2022).

Acknowledgments: We would like to thank two anonymous reviewers of *Languages* for their detailed and thoughtful comments and critiques on an earlier version of this paper, which we believe helped us sharpen and focus our framing and discussion.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Experiment Demonstration Links

The following links go to full working demonstrations of Experiments 2 and 3. To get past the onboarding form, fill in the mandatory fields with bogus information.

- Experiment 2: https://farm.pcibex.net/r/YfwLvt/ (accessed on 11 January 2022)
- Experiment 3: https://farm.pcibex.net/r/JQXOij/ (accessed on 11 January 2022)

Appendix B. Evidential Existentiality Norming Study

Appendix B.1. Participants

A total of 121 undergraduate students at UC Santa Cruz participated in the norming experiment for course credit—0 of these participants' data was not included in the analysis, 27 of which self-reported as non-native English speakers, and three of which met at least one of the exclusion criteria defined in (13). The data from ninety-one participants were included in the analysis. Participant age ranged from 18 to 33. The mean age was 20.

Appendix B.2. Materials and Methods

Thirty-six items were created, twelve of which were again reserved for the burn-in practice period. A sample item is provided in Table A1. The experiment included a single factor, Response, of which there were three levels: there existential, Evidential existential, and Transitive verb. These response types describe responses to polar questions inquiring about the existence of a human individual matching a particular description contained in a relative clause. The question was invariant within each item.

Table A1. Evidential existentiality norming study sample item.

	Sentence	Response Type
	Question: Is there anyone who can decode this script?	
a	Yeah, I'm sure there's someone who can decode it.	*There* existential
b	Yeah, I talked to someone who can decode it.	Evidential existential
c	Yeah, I criticized someone who can decode it.	Transitive verb

On a given trial, participants saw a polar question presented above one kind of response. The question–answer pair was formatted as a brief text-message thread (Figure 4). As in the other experiments, participants were instructed to choose a rating from a Likert-type scale. Here, they were instructed to rate how natural the response was to the answer.

Appendix B.3. Analysis

We fit a mixed-effects ordinal regression model with a cumulative link to the ratings data. A maximal random-effects structure was specified. Rating was set as the dependent variable, and Response was set as a fixed effect.

The Response factor was given Helmert contrast coding. This allowed for two comparisons: one direct comparison between the there *existential* and *Evidential existential* conditions and a comparison between the *Transitive verb* conditions and the mean of the two existential(-like) conditions.

Appendix B.4. Predictions

We expect a main effect of Response in which the *Transitive verb* responses are significantly different from the mean of the other two responses. Between the canonical There *existential* and *Evidential existential* response types, we expect the There *existentials* to be rated significantly higher (better).

Appendix B.5. Results

Mean ratings, standard deviation, and standard errors are reported for the norming study in Table A2. In the ordinal regression model, we observed a significant difference ($p = 0.031$) between the two existential response types as well as a significant difference ($p < 0.001$) between the Transitive verb response type and the mean of the existential response types. Model outputs are reported in Appendix E.2.

Table A2. Descriptive statistics for evidential existentiality norming study results. Mean is calculated on raw (non-z-scored) ratings.

Response	Mean Rating	SD	n	SE
There existential	4.8	1.4	728	0.052
Evidential existential	4.7	1.4	728	0.053
Transitive verb	3.2	1.6	728	0.061

Discussion

The predictions described above were borne out. There was a slight but significant difference between the naturalness of the there existential and Evidential existential in the context of the polar questions, but there was a marked difference between the naturalness of the transitive verb response type and the combination of the other two response types.

Appendix C. Burn-in Trial Results

The results of the burn-in trials are reported below for the evidential existentiality norming study, Experiment 2, and Experiment 3. Burn-in trials were not employed in Experiment 1.

Table A3. Descriptive statistics for evidential existentiality norming study burn-in trial results. A + or − in the rating cell (respectively) represents a higher or lower mean score for that condition in the burn-in trials than in the main trials. Mean is calculated on raw (non-z-scored) ratings.

Response	Mean Rating	SD	n	SE
There existential	4.8	1.3	364	0.070
Evidential existential	4.6 (−)	1.4	364	0.072
Transitive verb	3.2	1.6	364	0.081

Table A4. Descriptive statistics for Experiment 2 burn-in trial results. A + or − in the rating cell (respectively) represents a higher or lower mean score for that condition in the burn-in trials than in the main trials. Mean is calculated on raw (non-z-scored) ratings.

Length	Structure	Verb Type	Mean Rating	SD	n	SE
Short	Non-island	CP	4.15 (−)	1.35	161	0.106
Long	Non-island	CP	4.00 (+)	1.52	161	0.120
Short	Island	EE	3.99 (−)	1.33	161	0.105
Long	Island	EE	2.78 (−)	1.49	161	0.118
Short	Island	VT	3.66 (−)	1.43	161	0.113
Long	Island	VT	2.33 (−)	1.32	161	0.104

Table A5. Descriptive statistics for Experiment 3 burn-in trial results. A + or − in the rating cell (respectively) represents a higher or lower mean score for that condition in the burn-in trials than in the main trials. Mean is calculated on raw (non-z-scored) ratings.

Length	Structure	Verb Type	Mean Rating	SD	n	SE
Short	Non-island	CP	4.51 (−)	1.16	88	0.126
Long	Non-island	CP	4.06 (−)	1.20	88	0.130
Short	Island	EE	4.10 (−)	1.28	88	0.139
Long	Island	EE	2.51 (−)	1.21	88	0.132
Short	Island	VT	3.85 (−)	1.36	88	0.148
Long	Island	VT	2.10 (+)	1.09	88	0.119

Appendix D. DD Scores

Appendix D.1. Experiment 1 DD Scores

Table A6. DD score calculation table for Experiment 1, *Transitive object* conditions. The table displays the z-scored ratings for each condition and the intermediate differences used to calculate the DD score by item. Refer to Section 2 for a description of the procedure for calculating DD scores. The DD scores below are averaged for the *Transitive object* environment and presented in Figures 2 and 11.

Item	Env.	Non-Island, Short	Non-Island, Long	Island, Short	Islang, Long	D1	D2	DD
1	Tr. obj.	0.571	0.217	−0.299	−0.827	1.044	0.871	0.173
2	Tr. obj.	0.529	0.738	0.594	−0.522	1.260	−0.066	1.326
3	Tr. obj.	0.161	0.021	0.173	−0.754	0.775	−0.012	0.787
4	Tr. obj.	0.108	−0.683	−0.258	−0.198	−0.484	0.366	−0.851
5	Tr. obj.	0.161	−0.385	0.119	−1.477	1.092	0.043	1.049
6	Tr. obj.	0.605	0.119	0.728	0.302	−0.183	−0.123	−0.061
7	Tr. obj.	−0.604	0.926	0.297	−0.362	1.288	−0.901	2.189
8	Tr. obj.	1.057	−0.615	−0.782	−0.797	0.182	1.838	−1.656
9	Tr. obj.	0.608	0.267	−0.133	−0.227	0.494	0.741	−0.247
10	Tr. obj.	0.105	−0.458	−0.851	−0.873	0.416	0.956	−0.541
11	Tr. obj.	−0.256	−0.539	−0.436	−1.394	0.854	0.180	0.674
12	Tr. obj.	0.983	0.428	1.001	−0.186	0.614	−0.017	0.631
13	Tr. obj.	−0.250	0.197	0.110	−0.114	0.311	−0.361	0.672
14	Tr. obj.	−0.359	0.110	0.490	−0.856	0.966	−0.849	1.815
15	Tr. obj.	−0.028	0.421	0.533	−0.202	0.623	−0.560	1.183
16	Tr. obj.	−0.031	0.475	0.134	−0.669	1.144	−0.164	1.308
17	Tr. obj.	0.629	−0.528	0.173	−0.374	−0.155	0.457	−0.611
18	Tr. obj.	−1.009	−0.144	0.455	−0.952	0.808	−1.464	2.272
19	Tr. obj.	−0.109	0.633	0.259	−0.333	0.966	−0.369	1.335
20	Tr. obj.	0.065	−0.411	−0.214	−0.797	0.386	0.279	0.107
21	Tr. obj.	−0.073	−0.035	1.216	−0.089	0.054	−1.289	1.343
22	Tr. obj.	−0.170	0.552	0.141	−0.566	1.117	−0.311	1.429
23	Tr. obj.	−0.103	−0.527	−0.240	−0.387	−0.141	0.137	−0.278
24	Tr. obj.	0.568	0.725	0.213	−0.691	1.416	0.355	1.062
25	Tr. obj.	0.249	0.278	0.876	−0.428	0.706	−0.626	1.332
26	Tr. obj.	0.306	0.433	0.579	−0.547	0.980	−0.273	1.253
27	Tr. obj.	−0.237	0.316	0.383	−0.401	0.717	−0.620	1.337
28	Tr. obj.	0.282	−0.355	0.263	−0.132	−0.223	0.019	−0.242
29	Tr. obj.	−0.004	0.154	−0.186	−0.585	0.739	0.182	0.556
30	Tr. obj.	−0.244	−0.604	0.432	−1.065	0.462	−0.676	1.137
31	Tr. obj.	−0.322	0.703	0.817	−0.358	1.061	−1.139	2.201
32	Tr. obj.	0.197	0.302	−0.041	0.227	0.075	0.238	−0.162
33	Tr. obj.	0.549	0.282	−0.758	−0.401	0.683	1.307	−0.624
34	Tr. obj.	0.126	0.227	−0.147	−0.719	0.946	0.274	0.672
35	Tr. obj.	1.053	−0.239	−0.086	−0.440	0.201	1.140	−0.939
36	Tr. obj.	0.857	0.249	0.472	−0.920	1.170	0.385	0.785

Table A7. DD score calculation table for Experiment 1, *Predicate* conditions. The table displays the z-scored ratings for each condition and the intermediate differences used to calculate the DD score by item. Refer to Section 2 for a description of the procedure for calculating DD scores. The DD scores below are averaged for the *Predicate* environment and presented in Figures 2 and 11.

Item	Env.	Non-Island, Short	Non-Island, Long	Island, Short	Islang, Long	D1	D2	DD
1	Predicate	−0.057	0.735	0.345	−0.089	0.823	−0.402	1.225
2	Predicate	0.231	0.876	0.833	0.043	0.833	−0.602	1.435
3	Predicate	1.053	−0.239	0.019	−0.141	−0.098	1.034	−1.132
4	Predicate	−0.436	0.249	0.172	−0.039	0.288	−0.608	0.896
5	Predicate	0.557	−0.010	−0.406	−0.167	0.158	0.963	−0.806
6	Predicate	0.851	1.294	0.525	−0.439	1.733	0.325	1.408
7	Predicate	0.686	0.265	0.810	0.164	0.102	−0.124	0.225
8	Predicate	0.438	0.022	0.560	−0.492	0.514	−0.122	0.636
9	Predicate	−0.294	0.419	0.466	0.062	0.357	−0.761	1.118
10	Predicate	−0.102	−0.152	0.986	−0.272	0.120	−1.088	1.208
11	Predicate	−0.198	−0.349	−0.272	−0.797	0.448	0.073	0.375
12	Predicate	0.446	−0.544	0.101	−0.127	−0.417	0.345	−0.761
13	Predicate	1.026	0.092	−0.419	0.038	0.054	1.445	−1.390
14	Predicate	−0.239	−0.797	−0.298	−0.629	−0.168	0.059	−0.227
15	Predicate	0.561	−0.262	−0.110	0.093	−0.355	0.671	−1.026
16	Predicate	−0.239	0.029	1.001	−0.039	0.068	−1.239	1.308
17	Predicate	0.043	−0.514	0.548	−0.308	−0.206	−0.505	0.299
18	Predicate	0.452	−0.268	−0.119	−0.731	0.463	0.571	−0.108
19	Predicate	0.697	0.386	0.787	0.436	−0.049	−0.089	0.040
20	Predicate	0.281	−0.397	0.688	−0.314	−0.083	−0.407	0.324
21	Predicate	0.126	0.007	0.698	0.171	−0.163	−0.572	0.409
22	Predicate	0.445	−0.314	0.432	0.088	−0.402	0.013	−0.415
23	Predicate	−0.132	−0.607	−0.160	−0.524	−0.083	0.027	−0.111
24	Predicate	0.527	0.324	0.146	−0.127	0.452	0.381	0.070
25	Predicate	0.517	−0.060	0.231	−0.100	0.040	0.286	−0.246
26	Predicate	0.292	0.384	0.094	0.389	−0.006	0.197	−0.203
27	Predicate	0.546	0.453	0.095	0.551	−0.098	0.451	−0.549
28	Predicate	0.603	0.417	0.140	0.221	0.196	0.463	−0.267
29	Predicate	0.403	0.675	0.666	0.075	0.601	−0.263	0.864
30	Predicate	0.242	0.308	−0.114	−0.600	0.908	0.356	0.552
31	Predicate	0.876	0.195	0.785	−0.165	0.360	0.090	0.270
32	Predicate	0.177	−0.315	0.280	−0.314	0.000	−0.103	0.102
33	Predicate	0.136	−0.150	0.933	0.219	−0.369	−0.797	0.428
34	Predicate	0.247	−0.322	0.502	−0.171	−0.151	−0.255	0.104
35	Predicate	−0.229	−0.631	−0.662	−0.333	−0.297	0.433	−0.730
36	Predicate	0.326	0.367	0.011	−0.443	0.810	0.315	0.495

Table A8. DD score calculation table for Experiment 1, *Existential* conditions. The table displays the z-scored ratings for each condition and the intermediate differences used to calculate the DD score by item. Refer to Section 2 for a description of the procedure for calculating DD scores. The DD scores below are averaged for the *Existential* environment and presented in Figures 2 and 11.

Item	Env.	Non-Island, Short	Non-Island, Long	Island, Short	Islang, Long	D1	D2	DD
1	Existential	1.245	0.406	1.420	0.137	0.269	−0.174	0.444
2	Existential	1.453	0.814	1.187	−0.089	0.903	0.266	0.637
3	Existential	0.416	0.323	0.989	−0.225	0.549	−0.573	1.121
4	Existential	0.834	0.860	0.571	0.389	0.471	0.263	0.208
5	Existential	0.297	−0.607	0.399	−0.828	0.221	−0.101	0.322
6	Existential	1.520	1.053	0.533	0.725	0.328	0.986	−0.658
7	Existential	0.723	0.580	0.687	−0.081	0.661	0.036	0.625
8	Existential	0.857	0.043	0.067	0.319	−0.276	0.790	−1.066
9	Existential	0.725	0.628	1.085	0.058	0.570	−0.359	0.930
10	Existential	1.001	0.497	0.875	−0.708	1.205	0.126	1.079
11	Existential	−0.217	0.024	−0.260	−0.105	0.129	0.043	0.085
12	Existential	0.752	0.785	1.184	0.319	0.466	−0.433	0.899
13	Existential	0.943	0.134	1.273	0.396	−0.263	−0.330	0.067
14	Existential	0.860	−0.333	0.527	−0.305	−0.028	0.333	−0.361
15	Existential	1.308	0.597	0.968	0.733	−0.135	0.341	−0.476

Table A8. *Cont.*

Item	Env.	Non-Island, Short	Non-Island, Long	Island, Short	Islang, Long	D1	D2	DD
16	Existential	0.986	0.667	0.398	−0.089	0.756	0.588	0.167
17	Existential	0.549	0.227	1.216	0.200	0.027	−0.667	0.694
18	Existential	1.516	−0.266	0.533	−0.215	−0.050	0.983	−1.033
19	Existential	0.231	1.284	−0.288	0.448	0.836	0.519	0.316
20	Existential	0.683	0.725	0.778	0.181	0.544	−0.094	0.638
21	Existential	0.173	1.001	1.483	0.404	0.596	−1.309	1.906
22	Existential	0.778	1.065	0.683	−0.052	1.117	0.095	1.022
23	Existential	0.319	0.492	0.578	−0.276	0.767	−0.259	1.026
24	Existential	0.596	−0.294	1.168	−0.596	0.301	−0.572	0.874
25	Existential	1.221	1.029	0.667	0.396	0.632	0.554	0.079
26	Existential	1.012	0.644	1.057	−0.272	0.915	−0.044	0.960
27	Existential	1.879	1.187	0.989	−0.214	1.401	0.890	0.511
28	Existential	1.187	0.496	0.419	0.074	0.422	0.768	−0.346
29	Existential	1.310	0.419	0.231	0.327	0.092	1.080	−0.987
30	Existential	1.202	−0.266	0.707	0.197	−0.463	0.495	−0.958
31	Existential	0.723	0.453	0.365	0.582	−0.129	0.358	−0.487
32	Existential	0.984	0.711	0.778	0.666	0.045	0.206	−0.161
33	Existential	0.365	−0.130	0.916	−0.202	0.072	−0.550	0.622
34	Existential	1.001	0.308	1.082	−0.226	0.534	−0.081	0.615
35	Existential	−0.364	−0.514	0.019	−0.909	0.395	−0.383	0.779
36	Existential	1.569	−0.018	0.402	−0.322	0.304	1.167	−0.863

Appendix D.2. Experiment 2 DD Scores

Table A9. DD score calculation table for Experiment 2. The table displays the z-scored ratings for each condition and the intermediate differences used to calculate the DD score by item. Refer to Section 2 for a description of the procedure for calculating DD scores. The DD scores below are averaged for each verb and presented in Figures 7 and 10.

Item	Verb (EE)	Non-Island, Short	Non-Island, Long	Island, Short (EE)	Island, Long (EE)	D1 (EE)	D2 (EE)	DD (EE)
1	*find*	−0.038	−0.025	0.292	−0.237	0.212	−0.330	0.543
5	*find*	−0.014	0.018	−0.195	−0.592	0.611	0.180	0.430
9	*find*	−0.222	0.355	0.335	−0.835	1.190	−0.557	1.747
13	*find*	−0.064	−0.198	0.126	−0.662	0.464	−0.190	0.654
21	*find*	−0.213	−0.424	0.311	−0.915	0.491	−0.524	1.015
3	*hear of*	−0.718	0.000	−0.405	−0.900	0.900	−0.312	1.213
7	*hear of*	−0.214	−0.324	0.119	0.205	−0.529	−0.333	−0.197
11	*hear of*	−0.535	−0.518	−0.286	−0.407	−0.111	−0.249	0.137
15	*hear of*	−0.346	−0.255	0.367	−0.833	0.578	−0.713	1.291
19	*hear of*	−0.255	−0.779	−0.523	−0.366	−0.413	0.267	−0.680
23	*hear of*	−0.174	−0.568	−0.609	−1.113	0.546	0.435	0.110
2	*meet*	0.044	−0.651	0.219	−0.364	−0.287	−0.174	−0.113
6	*meet*	0.237	−0.345	−0.220	−1.049	0.704	0.457	0.247
10	*meet*	0.321	−0.411	0.008	−0.706	0.295	0.313	−0.018
14	*meet*	0.439	−0.202	0.355	−0.415	0.213	0.084	0.129
18	*meet*	−0.086	−0.359	0.038	−1.191	0.832	−0.124	0.957
22	*meet*	−0.376	−0.559	0.250	−0.448	−0.111	−0.626	0.516
4	*talk to*	0.264	−0.478	0.256	−0.653	0.175	0.008	0.167
8	*talk to*	0.047	−0.318	−0.102	−0.715	0.397	0.148	0.248
12	*talk to*	0.265	0.063	−0.054	−0.777	0.840	0.318	0.522
16	*talk to*	−0.240	−0.494	0.025	−0.553	0.059	−0.266	0.325
20	*talk to*	0.147	−0.283	−0.066	−0.511	0.228	0.213	0.015
24	*talk to*	0.105	−0.403	0.299	−0.566	0.164	−0.194	0.358

Table A9. *Cont.*

Item	Verb (VT)	Non-Island, Short	Non-Island, Long	Island, Short (VT)	Island, Long (VT)	D1 (VT)	D2 (VT)	DD (VT)
1	*slap*	−0.038	−0.025	−0.814	−1.600	1.576	0.775	0.800
5	*slap*	−0.014	0.018	−0.827	−2.116	2.134	0.813	1.321
9	*slap*	−0.222	0.355	−0.381	−1.366	1.720	0.159	1.562
13	*slap*	−0.064	−0.198	−0.684	−1.626	1.428	0.621	0.807
21	*slap*	−0.213	−0.424	−0.305	−1.241	0.817	0.092	0.725
3	*describe*	−0.718	0.000	−0.528	−0.864	0.864	−0.190	1.054
7	*describe*	−0.214	−0.324	0.062	−1.073	0.749	−0.276	1.025
11	*describe*	−0.535	−0.518	−0.175	−1.273	0.755	−0.359	1.114
15	*describe*	−0.346	−0.255	−0.020	−0.708	0.453	−0.327	0.779
19	*describe*	−0.255	−0.779	−0.850	−0.814	0.035	0.594	−0.559
23	*describe*	−0.174	−0.568	−0.308	−2.116	1.548	0.134	1.414
2	*imitate*	0.044	−0.651	0.286	−0.998	0.347	−0.241	0.588
6	*imitate*	0.237	−0.345	−0.338	−1.591	1.246	0.575	0.670
10	*imitate*	0.321	−0.411	0.413	−1.489	1.078	−0.092	1.170
14	*imitate*	0.439	−0.202	−0.041	−0.889	0.687	0.480	0.207
18	*imitate*	−0.086	−0.359	−0.244	−1.376	1.017	0.158	0.859
22	*imitate*	−0.376	−0.559	−0.392	−0.954	0.396	0.016	0.380
4	*criticize*	0.264	−0.478	−1.085	−1.419	0.942	1.349	−0.407
8	*criticize*	0.047	−0.318	−0.197	−0.845	0.526	0.244	0.282
12	*criticize*	0.265	0.063	−0.286	−1.476	1.539	0.551	0.988
16	*criticize*	−0.240	−0.494	−0.849	−1.229	0.735	0.609	0.126
20	*criticize*	0.147	−0.283	−0.423	−1.088	0.804	0.569	0.235
24	*criticize*	0.105	−0.403	−0.556	−1.046	0.643	0.661	−0.018

Appendix D.3. Experiment 3 DD Scores

Table A10. DD score calculation table for Experiment 3. The table displays the z-scored ratings for each condition and the intermediate differences used to calculate the DD score by item. Refer to Section 2 for a description of the procedure for calculating DD scores. The DD scores below are averaged for each verb and presented in Figures 9 and 10.

Item	Verb (EE)	Non-Island, Short	Non-Island, Long	Island, Short (EE)	Island, Long (EE)	D1 (EE)	D2 (EE)	DD (EE)
1	*find*	0.090	0.013	0.220	−0.208	0.221	−0.130	0.350
5	*find*	0.147	−0.010	0.332	−1.064	1.054	−0.185	1.239
9	*find*	−0.091	−0.049	0.003	−0.592	0.543	−0.093	0.637
13	*find*	0.182	0.585	−0.317	−1.041	1.626	0.499	1.127
21	*find*	−0.147	−0.020	0.042	−0.769	0.750	−0.189	0.938
3	*hear of*	−0.095	0.284	−0.324	−1.049	1.333	0.229	1.104
7	*hear of*	0.347	0.082	0.074	−0.452	0.534	0.273	0.261
11	*hear of*	−0.222	−0.099	−0.115	−1.040	0.941	−0.107	1.048
15	*hear of*	−0.266	0.073	−0.031	−0.213	0.285	−0.235	0.521
19	*hear of*	−0.200	−0.039	0.055	−1.415	1.375	−0.255	1.630
23	*hear of*	0.037	−0.345	−0.251	−1.080	0.734	0.287	0.447
2	*meet*	0.365	0.088	0.563	−0.568	0.656	−0.198	0.854
6	*meet*	0.190	0.166	0.211	−0.794	0.960	−0.021	0.981
10	*meet*	0.377	0.127	0.131	−0.810	0.936	0.246	0.690
14	*meet*	0.591	0.106	0.300	−0.636	0.741	0.291	0.450
18	*meet*	0.273	0.255	−0.132	−1.462	1.716	0.405	1.312
22	*meet*	0.432	−0.170	0.014	−0.957	0.787	0.418	0.369
4	*talk to*	0.506	−0.199	−0.246	−1.324	1.125	0.752	0.373
8	*talk to*	0.492	−0.431	−0.001	−0.932	0.501	0.493	0.009
12	*talk to*	0.001	0.052	−0.173	−0.689	0.741	0.174	0.567
16	*talk to*	0.580	−0.424	−0.305	−0.918	0.495	0.885	−0.390
20	*talk to*	0.347	−0.039	0.009	−0.630	0.591	0.338	0.253
24	*talk to*	0.267	−0.113	0.420	−0.516	0.402	−0.152	0.555

Table A10. *Cont.*

Item	Verb (VT)	Non-Island, Short	Non-Island, Long	Island, Short (VT)	Island, Long (VT)	D1 (VT)	D2 (VT)	DD (VT)
1	*slap*	0.090	0.013	−0.400	−1.381	1.394	0.490	0.904
5	*slap*	0.147	−0.010	−0.235	−1.566	1.556	0.382	1.174
9	*slap*	−0.091	−0.049	−0.863	−1.161	1.112	0.772	0.340
13	*slap*	0.182	0.585	−0.979	−1.139	1.724	1.161	0.563
21	*slap*	−0.147	−0.020	−0.297	−1.724	1.704	0.150	1.554
3	*describe*	−0.095	0.284	0.081	−0.750	1.034	−0.176	1.209
7	*describe*	0.347	0.082	0.103	−1.149	1.231	0.245	0.986
11	*describe*	−0.222	−0.099	−0.120	−0.829	0.730	−0.102	0.832
15	*describe*	−0.266	0.073	0.067	−0.964	1.036	−0.333	1.369
19	*describe*	−0.200	−0.039	−0.255	−1.078	1.038	0.056	0.983
23	*describe*	0.037	−0.345	−0.494	−0.940	0.595	0.531	0.064
2	*imitate*	0.365	0.088	−0.024	−1.041	1.129	0.388	0.741
6	*imitate*	0.190	0.166	0.082	−2.110	2.277	0.108	2.169
10	*imitate*	0.377	0.127	−0.168	−1.239	1.366	0.545	0.821
14	*imitate*	0.591	0.106	−0.609	−1.366	1.472	1.200	0.272
18	*imitate*	0.273	0.255	0.027	−1.439	1.694	0.245	1.448
22	*imitate*	0.432	−0.170	−0.205	−0.833	0.663	0.637	0.026
4	*criticize*	0.506	−0.199	−0.488	−1.891	1.691	0.994	0.697
8	*criticize*	0.492	−0.431	0.359	−1.129	0.699	0.133	0.566
12	*criticize*	0.001	0.052	−0.324	−1.197	1.249	0.325	0.924
16	*criticize*	0.580	−0.424	0.488	−0.853	0.429	0.092	0.338
20	*criticize*	0.347	−0.039	−0.076	−1.444	1.405	0.424	0.982
24	*criticize*	0.267	−0.113	0.402	−0.964	0.851	−0.135	0.986

Appendix E. Ordinal Regression Model Outputs

Appendix E.1. Experiment 1 Ordinal Regression Model Output

(14) Formula for Experiment 1 mixed-effects ordinal regression

```
rating ~ environment * structure * length +
(1 + environment * structure * length | subject) +
(1 + environment * structure * length | item)
```

Table A11. Contrasts for Experiment 1 Environment factor (Helmert coding).

Factor Level	Pred_Exist	PredExist_Object
Transitive object	0	2
Predicate	1	−1
Existential	−1	−1

Table A12. Contrasts for Experiment 1 Structure factor.

Factor Level	Contrast Value
Non-island	−0.5
Island	0.5

Table A13. Contrasts for Experiment 1 Length factor.

Factor Level	Contrast Value
Short	−0.5
Long	0.5

Table A14. Coefficient estimates for Experiment 1 mixed-effects model.

Effect	β	SE	z	p
Pred_Exist	−0.7346	0.1046	−7.03	∼0
PredExist_Object	−0.4327	0.0626	−6.92	∼0
Structure	0.1411	0.1411	5.65	∼0
Length	1.7243	0.2470	6.98	∼0
Pred_Exist × Structure	−0.3824	0.1967	−1.94	0.052
PredExist_Object × Structure	0.0966	0.1080	0.89	0.371
Pred_Exist × Length	−0.4487	0.1806	−2.49	0.013
PredExist_Object × Length	−0.0781	0.0959	−0.81	0.415
Structure × Length	−1.2935	0.2872	−4.50	∼0
Pred_Exist × Structure × Length	0.2203	0.3764	0.59	0.558
PredExist_Object × Structure × Length	−0.4544	0.2109	−2.15	0.031

Appendix E.2. Evidential Existentiality Norming Study Ordinal Regression Model Output

(15) Formula for evidential existentiality norming study mixed-effects ordinal regression

```
rating ~ response +
(1 + response | subject) +
(1 + response | item)
```

Table A15. Contrasts for evidential existentiality norming study Response factor (Helmert coding).

Factor Level	Exist_EE	ExistEE_VT
There existential	−1	−1
Evidential existential	1	−1
Transitive verb	0	2

Table A16. Coefficient estimates for evidential existentiality norming study mixed-effects model.

Effect	β	SE	z	p
ResponseExist_EE	−0.4308	0.1993	−2.162	0.0306
ResponseExistEE_VT	−3.1935	0.4089	−7.811	∼0

Appendix E.3. Experiment 2 Ordinal Regression Model Output

(16) Formula for Experiment 2 mixed-effects ordinal regression

```
rating ~ length * verbtype +
(1 + length * verbtype | subject) +
(1 + length * verbtype | item)
```

Table A17. Contrasts for Experiment 2 Verb type factor.

Factor Level	CP–EE	CP–VT
CP	0	0
EE	1	0
VT	0	1

Table A18. Contrasts for Experiment 2 Length factor.

Factor Level	Contrast Value
Short	−0.5
Long	0.5

Table A19. Coefficient estimates for Experiment 2 mixed-effects model.

Effect	β	SE	z	p
Length	−1.1022	0.4796	−2.298	0.022
VerbtypeCP_EE	−0.3980	0.2150	−1.851	0.064
VerbtypeCP_VT	−2.1885	0.3123	−7.008	∼0
Length × VerbtypeCP_EE	−1.3109	0.3828	−3.425	∼0
Length × VerbtypeCP_VT	−2.3789	0.4667	−5.097	∼0

Appendix E.4. Experiment 3 Ordinal Regression Model Output

(17) Formula for Experiment 3 mixed-effects ordinal regression

```
rating ~ length * verbtype +
(1 + length * verbtype | subject) +
(1 + length * verbtype | item)
```

See Tables A17 and A18 for the contrasts assigned to Verb type and Length, respectively, which were the same as in Experiment 2.

Table A20. Coefficient estimates for Experiment 3 mixed-effects model.

Effect	β	SE	z	p
Length	−0.7420	0.2792	−2.658	∼0
VerbtypeCP_EE	−1.6244	0.2231	−7.282	∼0
VerbtypeCP_VT	−2.7571	0.3601	−7.656	∼0
Length × VerbtypeCP_EE	−2.1150	0.3902	−5.421	∼0
Length × VerbtypeCP_VT	−3.1014	0.5582	−5.556	∼0

Appendix E.5. Combined Experiment 2 and 3 Ordinal Regression Model Output

(18) Formula for combined Experiment 2–Experiment 3 mixed-effects ordinal regression

```
rating ~ length * verbtype * experiment +
(1 + length * env | subject) +
(1 + length * env | item)
```

Table A21. Contrasts for Experiment factor.

Factor Level	Contrast Value
Exp 2	−0.5
Exp 3	0.5

See Tables A17 and A18 for the contrasts assigned to Verb type and Length, respectively, which were the same as in Experiment 2.

Table A22. Coefficient estimates for combined Experiment 2 and 3 mixed-effects model.

Effect	β	SE	z	p
Length	−0.8867	0.2573	−3.445	0.001
VerbtypeCP_EE	−1.0012	0.1631	−6.137	∼0
VerbtypeCP_VT	−2.4287	0.2572	−9.443	∼0
Experiment	1.0407	0.3967	2.624	0.009
Length × VerbtypeCP_EE	−1.7145	0.2516	−6.813	∼0
Length × VerbtypeCP_VT	−2.7204	0.3519	−7.731	∼0
Length × Experiment	0.1127	0.4802	0.235	0.815
VerbtypeCP_EE × Experiment	−1.2659	0.2705	−4.680	∼0
VerbtypeCP_VT × Experiment	−0.6230	0.3486	−1.787	0.074
Length × VerbtypeCP_EE × Experiment	−0.7531	0.4698	−1.603	0.109
Length × VerbtypeCP_VT × Experiment	−0.7238	0.6113	−1.184	0.236

Notes

[1] Two of the studies reported here were conducted in the course of the research for the first author's unpublished dissertation (Vincent 2021): Experiment 1 was conducted in 2018 and is presented in Vincent (2021) as Experiment 3; the norming study discussed in Section 3.2 was conducted in 2020 and is presented in Vincent (2021) as Experiment 6.

[2] In an Evidential Existential, the speaker asserts (or denies) existence of the entity denoted by the RC in a first-person statement whose predicate indicates the source of evidence for the existential claim: *I know*, *I saw*, *I heard of* (Rubovitz-Mann 2000, 2012).

[3] Here and in the remainder of this paper, we use the term *permeable* to describe a relative clause that permits extraction of a phrase from within its boundaries to a position outside of its boundaries.

[4] The full list of predicates is *kjenne til* ("be acquainted with"), *snakke med* ("speak with"), *vite om* ("know about/of"), and *møte* ("meet") (Dave Kush, p.c.). It is worth observing that these could each plausibly have an evidential existential use. In separate experiments, Kush et al. (2018) used *wh*-phrases that were bare (Exp. 1–2: *hvem*/*hva*, 'who'/'what') or complex (Exp. 3: *hvilken regissør*/*film*, 'which director/film'); they found a large positive DD score even when the *wh*-phrase was complex.

[5] For a relevant critique of the factorial definition of islands that was brought to our attention by an anonymous reviewer, see Kim (2021). See Vincent (2021, pp. 67–71) for discussion of some of the challenges associated with minimizing confounding factors in the factorial definition of islands as it is used in the current experiment.

[6] As the reader will discover in the sections about Experiments 2 and 3, there is some variation in the effect that different transitive verbs have on relative clause permeability. This is not something that we controlled for in Experiment 1, and we assume that the sample of transitive verbs used for the *Transitive object* conditions on average represent the effect of an "average" transitive verb on relative clause permeability.

[7] An anonymous reviewer has asked why the exclusion criteria defined here are somewhat more aggressive than those for Experiment 1. The difference is a reflection of the substantial amount of time that passed between data collection for Experiment 1 and Experiments 2 and 3. Given how conservative the exclusion criterion was for Experiment 1, we believe that the change was a reasonable upgrade that is more capable of excluding data from participants who did not pay attention to the task.

[8] Due to the Latin Square counterbalancing employed, this resulted in the first twelve experimental trials each participant rated being set aside before analysis.

[9] One of the challenges associated with minimizing confounding factors in the factorial definition of islands, especially designs like ours which compare extraction from a complement clause with extraction from a DP, is that lexical differences are required across the *Non-island* and *Island* conditions; (see, e.g., Kim 2021, which was brought to our attention by an anonymous reviewer). This concern is perhaps mitigated to an extent if verbs are selected which accept either a CP complement or a DP complement (as in the *Transitive object* conditions of Experiment 1), but this results in severe limitations to item construction and raises non-trivial questions of its own, such as how ratings for a sentence containing a verb that *can* accept a CP complement are affected by one of the uses of that verb being more or less dominant.

[10] This item contained *find* in the EE conditions and *slap* in the VT conditions. The reader will notice that as a consequence of this, there is one less item in the list of DD scores by item in Appendices D.2 and D.3.

[11] Other factors may of course conspire to exclude extraction, even in these contexts; see Sichel (2018) for the significance of the structure of the containing RC, raising vs. head-external.

References

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204: 104293. [CrossRef] [PubMed]

Allwood, Jens. 1976. The Complex NP Constraint as a non-universal rule and some semantic factors influencing the acceptability of Swedish sentences which violate the CNPC. *University of Massachusetts Occasional Papers in Linguistics* 2: 1.

Allwood, Jens. 1982. The Complex NP Constraint in Swedish. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabet Engdahl and Eva Ejerhed. Stockholm: Almqvist & Wiksell International, vol. 43, pp. 15–32.

Almeida, Diogo. 2014. Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *The Journal of the Brazilian Linguistics Association* 13: 55. [CrossRef]

Ambridge, Ben, and Adele E. Goldberg. 2008. The Island Status of Clausal Complements: Evidence in Favor of an Information Structure Explanation. *Cognitive Linguistics* 19: 349–81. [CrossRef]

Bianchi, Valentina, and Cristiano Chesi. 2014. Subject Islands, Reconstruction, and the Flow of Computation. *Linguistic Inquiry* 45: 525–69. [CrossRef]

Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Edited by Robert Freidin, Carlos P. Otero and Maria L. Zubizarreta. Cambridge, MA: The MIT Press, vol. 45.

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2022. The island is still there: Experimental evidence for the inescapability of relative clauses in English. *Studia Linguistica*. [CrossRef]

Chung, Sandra, and James McCloskey. 1983. On the Interpretation of Certain Island Facts in GPSG. *Linguistic Inquiry* 14: 704–13.

Cinque, Guglielmo. 2010. On a Selective 'Violation' of the Complex NP Constraint. In *Structure Preserved: Studies in Syntax for Jan Koster*. Edited by Jan-Wouter Zwart and Mark de Vries. Amsterdam and Philadelphia: John Benjamins Publishing Company, vol. 164, pp. 81–89.
Davies, William D., and Stanley Dubinsky. 2003. On extraction from NPs. *Natural Language & Linguistic Theory* 21: 1–37. [CrossRef]
Deal, Amy Rose. 2009. The origin and content of expletives: Evidence from "selection". *Syntax* 12: 285–323. [CrossRef]
Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: The MIT Press.
Dillon, Brian, and Matthew W. Wagers. 2021. Approaching gradience in acceptability with the tools of signal detection theory. In *The Cambridge Handbook of Experimental Syntax*. Edited by Grant Goodall. Cambridge: Cambridge University Press, pp. 62–96.
Engdahl, Elisabet. 1982. Restrictions on Unbounded Dependencies in Swedish. In *Readings on Unbounded Dependencies in Scandinavian Languages, vol. 43 of Umeå Studies in the Humanities*. Edited by Elisabet Engdahl and Eva Ejerhed. Stockholm: Almqvist & Wiksell International, pp. 151–74.
Engdahl, Elisabet. 1997. Relative Clause Extractions in Context. *Working Papers in Scandinavian Syntax* 60: 51–79.
Erteschik-Shir, Nomi. 1973. On the Nature of Island Constraints. Ph.D. thesis, MIT, Cambridge, MA, USA.
Erteschik-Shir, Nomi. 1982. Extractability in Danish and the Pragmatic Principle of Dominance. In *Readings on Unbounded Dependencies in Scandinavian Languages, vol. 43 of Umeå Studies in the Humanities*. Edited by Elisabet Engdahl and Eva Ejerhed. Stockholm: Almqvist & Wiksell International, pp. 175–91.
Erteschik-Shir, Nomi, and Shalom Lappin. 1979. Dominance and the functional explanation of island phenomena. *Theoretical linguistics* 6: 41–86. [CrossRef]
Fiengo, Robert, and James Higginbotham. 1981. Opacity in NP. *Linguistic Analysis* 7: 395–421.
Gallego, Angel, and Juan Uriagereka. 2006. Sub-extraction from Subjects: A Phase Theory Account. In *Romance Linguistics 2006*. Edited by José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez and Maria José Cabrera. Amsterdam: John Benjamins, pp. 149–62.
Gallego, Angel, and Juan Uriagereka. 2007. Conditions on Sub-extraction. In *Coreference, Modality, and Focus*. Edited by Luis Eguren and Olga Fernández-Soriano. Amsterdam: John Benjamins, pp. 45–70.
Hofmeister, Philip, and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415. [CrossRef] [PubMed]
Keshev, Maayan, and Aya Meltzer-Asscher. 2018. A processing-based account of subliminal wh-island effects. *Natural Language & Linguistic Theory* 37: 621–57. [CrossRef]
Kim, Ilkyu. 2021. A Note on the Factorial Definition of Island Effects. *Lanaguage Research* 57: 211–23. [CrossRef]
Kluender, Robert. 1992. Deriving Island Constraints from Principles of Predication. In *Island Constraints: Theory, Acquisition and Processing*. Edited by Helen Goodluck and Michael Rochemont. Dordrecht: Kluwer Academic Publishers, vol. 15, pp. 223–58.
Kluender, Robert, and Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language & Cognitive Processes* 8: 573–633. [CrossRef]
Kuno, Susumu. 1976. Subject, Theme, and the Speaker's Empathy—A Reexamination of Relativization Phenomena. In *Subject and Topic*. Edited by Charles N. Li. New York: Academic Press, Inc., pp. 417–44.
Kuno, Susumu. 1987. *Functional Syntax: Anaphora, Discourse, and Empathy*. Chicago: University of Chicago Press.
Kush, Dave, Akira Omaki, and Norbert Hornstein. 2013. Microvariation in islands? In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 239–64.
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language & Linguistic Theory* 36: 743–79. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95: 393–420. [CrossRef]
Lindahl, Filippa. 2014. Relative Clauses are not always strong islands. *Working Papers in Scandinavian Syntax* 93: 1–25.
Lindahl, Filippa. 2017. Extraction from Relative Clauses in Swedish. Ph.D. thesis, University of Gothenburg, Gothenburg, Sweden.
Loftus, Geoffrey R. 1978. On interpretation of interactions. *Memory & Cognition* 6: 312–19. [CrossRef]
Löwenadler, John. 2015. Relative Clause Extraction: Pragmatic Dominance, Processing Complexity and the Nature of Crosslinguistic Variation. *Nordic Journal of Linguistics* 38: 37–65. [CrossRef]
Mahajan, Anoop. 1992. The Specificity Condition and the CED. *Linguistic Inquiry* 23: 510–16.
Maling, Joan, and Annie Zaenen. 1982. A phrase structure account of Scandinavian extraction phenomena. In *The Nature of Syntactic Representation, vol. 15 of Synthese Language Library*. Edited by Pauline Jacobson and Geoffrey K. Pullum. Dordrecht: D. Reidel Publishing Co., pp. 229–82
McCawley, James D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53: 99–149. [CrossRef]
McNally, Louise. 2008. DP-internal "only", Amount Relatives, and Relatives out of Existentials. *Linguistic Inquiry* 39: 161–69. [CrossRef]
Milsark, Gary Lee. 1974. Existential Sentences in English. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Moro, Andrea. 1997. *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge: Cambridge University Press.
Müller, Christiane. 2014. Swedish Relative Clause Extractions: The Small Clause Hypothesis. Master's thesis, Lund University, Lund, Sweden.
Müller, Christiane. 2015. Against the Small Clause Hypothesis: Evidence from Swedish relative clause extractions. *Nordic Journal of Linguistics* 38: 67–92. [CrossRef]

Nyvad, Anne Mette, Ken Ramshøj Christensen, and Sten Vikner. 2017. CP-recursion in Danish: A cP/CP-analysis. *The Linguistic Review* 34: 3. [CrossRef]
Prolific. 2022. Prolific (December 2021). Available online: https://www.prolific.co (accessed on 27 December 2021).
Rizzi, Luigi. 2004. Locality and Left Periphery. In *Structures and Beyond: The Cartography of Syntactic Structures*. Edited by Adriana Belleti. Oxford: Oxford University Press, vol. 3, pp. 223–51.
Ross, John Robert. 1967. Constraints on Variables in Syntax. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Rotello, Caren M., Evan Heit, and Chad Dubé. 2015. When more data steer us wrong: Replications with the wrong dependent measure perpetuate erroneous conclusions. Psychonomic Bulletin & Review 22: 944–54. [CrossRef]
Rubovitz-Mann, Talia. 2000. Extractions from Relative Clauses: An Information-Structural Account. Ph.D. thesis, Hebrew University of Jerusalem, Jerusalem, Israel.
Rubovitz-Mann, Talia. 2012. *Evidential-Existentials: An Information Structure Account of Extraction from Relative Clauses*. Sunnyvale: Lambert Academic Publishing.
Sichel, Ivy. 2018. Anatomy of a Counterexample: Extraction from relative clauses. *Linguistic Inquiry* 49: 335–78. [CrossRef]
Sprouse, Jon. 2007. A Program for Experimental Syntax: Finding the Relationship between Acceptability and Grammatical Knowledge. Ph.D. thesis, University of Maryland, College Park, MD, USA.
Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43: 155–67. [CrossRef]
Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34: 307–44. [CrossRef]
Sprouse, Jon, Matthew W. Wagers, and Colin Phillips. 2012. A test of the relation between working memory capacity and syntactic island effects. *Language* 88: 82–123. [CrossRef]
Sprouse, Jon, Shin Fukuda, Hajime Ono, and Robert Kluender. 2011. Reverse Island Effects and the Backward Search for a Licensor in Multiple Wh-Questions. *Syntax* 14: 179–203. [CrossRef]
Stepanov, Arthur. 2001. Late Adjunction and Minimalist Phrase Structure. *Syntax* 4: 94–125. [CrossRef]
Stepanov, Arthur, Mušič Manca, and Penka Stateva. 2018. Two (non-)islands in Slovenian: A study in experimental syntax. *Linguistics* 56: 435–76. [CrossRef]
Takahashi, Daiko. 1994. Minimality of Movement. Ph.D. thesis, University of Connecticut, Storrs, CT, USA.
Taraldsen, Knut Tarald. 1981. The Theoretical Interpretation of a Class of "Marked" Extractions. In *Theory of Markedness in Generative Grammar: Proceedings of the 1979 GLOW Conference*. Edited by Adriana Belletti, Luciana Brandi and Luigi Rizzi. Pisa: Scuola Normale Superiore di Pisa, pp. 475–516.
Taraldsen, Knut Tarald. 1982. Extraction from relative clauses in Norwegian. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabet Engdahl, Eva Ejerhed and Per-Göran Råberg. Stockholm: Almqvist & Wiksell International, vol. 43, pp. 205–21.
Tucker, Matthew A., Ali Idrissi, Jon Sprouse, and Diogo Almeida. 2019. Resumption Ameliorates Different Islands Differentially: Acceptability Data from Modern Standard Arabic. In *Perspectives on Arabic Linguistics XXX: Papers from the Annual Symposia on Arabic Linguistics, Stony Brook, New York, 2016 and Norman, Oklahoma, 2017*. Edited by Amel Khalfaoui and Matthew A. Tucker. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 159–94. [CrossRef]
Uriagereka, Juan. 1999. Multiple Spell-out. In *Working Minimalism*. Edited by Samuel D. Epstein and Norbert Hornstein. Cambridge, MA: MIT Press, pp. 251–82.
Vincent, Jake. 2021. Extraction from Relative Clauses: An Experimental Investigation into Variable Island Effects in English—Or—This Is a Dissertation That We Really Needed to Find Someone Who'd Write. Ph.D. thesis, University of California, Santa Cruz, Santa Cruz, CA, USA. Available online: https://escholarship.org/uc/item/5k08q709 (accessed on 27 December 2021).

languages

MDPI

Article

Comparing Island Effects for Different Dependency Types in Norwegian

Anastasia Kobzeva [1,*], Charlotte Sant [2], Parker T. Robbins [3], Myrte Vos [2], Terje Lohndal [1,2] and Dave Kush [1,3]

[1] Department of Language and Literature, Norwegian University of Science and Technology (NTNU), N-7491 Trondheim, Norway; terje.lohndal@ntnu.no (T.L.); dave.kush@utoronto.ca (D.K.)
[2] Department of Language and Culture, UiT The Arctic University of Norway, N-9037 Tromsø, Norway; charlotte.sant@uit.no (C.S.); myrte.t.vos@uit.no (M.V.)
[3] Department of Language Studies, University of Toronto, 1265 Military Trail, Toronto, ON M1C 1A4, Canada; parker.robbins@mail.utoronto.ca
* Correspondence: anastasia.kobzeva@ntnu.no

Abstract: Recent research suggests that island effects may vary as a function of dependency type, potentially challenging accounts that treat island effects as reflecting uniform constraints on all filler-gap dependency formation. Some authors argue that cross-dependency variation is more readily accounted for by discourse-functional constraints that take into account the *discourse status* of both the filler and the constituent containing the gap. We ran a judgment study that tested the acceptability of *wh*-extraction and relativization from nominal subjects, embedded questions (EQs), conditional adjuncts, and existential relative clauses (RCs) in Norwegian. The study had two goals: (i) to systematically investigate cross-dependency variation from various constituent types and (ii) to evaluate the results against the predictions of the FOCUS BACKGROUND CONFLICT constraint (FBCC). Overall we find some evidence for cross-dependency differences across extraction environments. Most notably *wh*-extraction from EQs and conditional adjuncts yields small but statistically significant island effects, but relativization does not. The differential island effects are potentially consistent with the predictions of the FBCC, but we discuss challenges the FBCC faces in explaining finer-grained judgment patterns.

Keywords: island constraints; experimental syntax; *wh*-questions; relative clauses; Norwegian

Citation: Kobzeva, Anastasia, Charlotte Sant, Parker T. Robbins, Myrte Vos, Terje Lohndal, and Dave Kush. 2022. Comparing Island Effects for Different Dependency Types in Norwegian. *Languages* 7: 195. https://doi.org/10.3390/languages7030195

Academic Editors: Anne Mette Nyvad and Ken Ramshøj Christensen

Received: 16 February 2022
Accepted: 20 July 2022
Published: 25 July 2022

1. Introduction

Natural languages can form *filler-gap dependencies*, which establish a relationship between a moved element (the *filler*) and a *gap* in its base syntactic position (i.e., where the filler is ultimately interpreted).[1] In *wh*-questions such as (1-a), the filler *wh*-phrase *which book* is linked to a *gap* contained within the complement clause. Relative clauses (RCs) such as (1-b) are also filler-gap dependencies, where the head of the RC is the filler that is linked to a gap.

(1) a. Which book$_i$ did Anna say [that Brian had read $\_\__i$]?
b. That is the book$_i$ which Anna said [that Brian had read $\_\__i$].

Filler-gap dependencies can in principle cross an arbitrary linear and structural distance (Chomsky1973, 1977), as illustrated in (2):

(2) a. Which book$_i$ did Anna say [that Sunniva thought [that Kristin believed [that Brian had read $\_\__i$]]]?
b. That was the book which$_i$ Anna said [that Sunniva thought [that Kristin believed [that Brian had read $\_\__i$]]].

Although long-distance filler-gap dependencies are possible, it has been known at least since Ross (1967) that trying to relate a filler to a gap inside specific constituents

leads to unacceptability. These domains are called *islands*. Several constituent types have been identified as islands, including subject phrases (nominal or clausal), certain adjuncts, embedded questions (EQs), and relative clauses (RCs) (Chomsky 1973, 1977; Huang 1982; Ross 1967; Stepanov 2007). Examples of these island types are given in (3).

(3) a. **Subject**
*Which boy$_i$ did you think that [the mother of __$_i$] was interesting?
b. **Adjunct**
*Which boy$_i$ did Christian talk to Odd [after Anne yelled at __$_i$]?
c. **Embedded Question**
*Which boy did Odd remember [what __$_i$ was called]?
d. **Relative Clause**
*Which cake$_i$ did you meet the woman [who made __$_i$]?

Following recent experimental work, we label the unacceptability that arises with such filler-gap dependencies *island effects* (Sprouse 2007; Sprouse et al. 2012, 2016).

Since the discovery of island effects, researchers have been interested in figuring out why they arise. A dominant tradition has sought to explain island effects as arising from universal syntactic conditions on A'-movement operations (Chomsky 1973, 1977, 1986, 2000; Cinque 1990; Huang 1982).[2] The traditional syntactic approach predicts, all else equal, that island effects should be observed with all dependencies that are derived via A'-movement, such as *wh*-movement and relativization (Chomsky 1977; Schütze et al. 2015).

An alternate functionalist tradition attributes island effects to discourse-pragmatic factors grounded in the information status of different elements in a sentence (e.g., Erteschik-Shir 1973; Goldberg 2006; Kuno 1987; Van Valin 1995). Particulars of individual accounts differ, but most employ the distinction between items that are in *focus* (those that correspond to or request new information) and those that are *backgrounded* (e.g., items that are given or *discourse-old*). The underlying intuition behind many of these accounts is that island effects arise when prominent or focused items are linked to gaps in backgrounded constituents. For example, Goldberg (2006) proposed that all filler-gap dependencies place the filler in a discourse prominent position, which is incompatible with gaps that fall inside backgrounded constituents. As a result the account also predicts that backgrounded constituents are islands for filler-gap dependencies.

(4) Backgrounded Constituents are Islands (BCI)
Backgrounded constituents may not serve as gaps in filler-gap constructions.
(Goldberg 2006, p. 135)

In apparent contradiction to the predictions of both traditional syntactic accounts and discourse-based accounts such as Goldberg (2006), recent experimental research suggests that certain island effects may vary as a function of A'-dependency type (Abeillé et al. 2020; Bondevik et al. 2021; Kush et al. 2018 2019; Sprouse et al. 2016). The extent of cross-dependency variation is, however, not well established. Moreover, the conclusion that different dependency types yield different island effects has been made based on comparison across experiments. Few studies have directly compared different dependency types within a single experiment.

The first goal of this paper, therefore, is to more systematically map the empirical landscape in one language, Norwegian, through a side-by-side comparison of island effects with *wh*- and RC-dependencies.

The second goal of the paper is to evaluate our results against a new discourse-based account of island effects, the Focus Background Conflict constraint (henceforth FBCC) put forward by Abeillé et al. (2020), which was developed specifically with the goal of accounting for cross-dependency variation in island effects. To keep the size of the paper manageable, we focus primarily on the FBCC and do not attempt to exhaustively cover how prior syntactic and discourse-based approaches could or would account for our findings.

Before we present our experiment and the results, the remainder of the introduction reviews the FBCC and provides some relevant background on islands in Norwegian.

1.1. The Focus-Background Conflict Constraint

Abeillé et al. (2020) proposed a new discourse-based constraint intended to account for island effects:

(5) FOCUS-BACKGROUND CONFLICT CONSTRAINT (FBCC)
A focused element should not be part of a backgrounded constituent.
(Abeillé et al. 2020, ex. 8)

According to the FBCC, whenever a focused filler is associated with a gap inside a backgrounded constituent, a clash in discourse-status occurs, causing the sentence to be infelicitous (rather than syntactically ill-formed). This infelicity results in a decrease in acceptability.[3] The FBCC links islandhood to backgroundedness, but unlike Goldberg's BCI (4), the FBCC is stated in such a way that it does not uniformly treat backgrounded constituents as islands for all filler-gap dependencies. Instead, the FBCC holds that backgrounded constituents are only islands for dependencies where the filler is focalized. *Wh*-dependencies put the questioned element into focus (Jackendoff 1972) by seeking new information, so *wh*-extraction from a backgrounded constituent is predicted to be unacceptable. RC-dependencies, however, do not place the filler—the head of the RC—into focus, because the function of a standard RC is to add information to a *given* entity. Therefore, the FBCC predicts that RC-dependencies into backgrounded constituents should be felicitous.

Abeillé and colleagues tested the predictions of the FBCC by investigating the acceptability of *wh*- and RC-dependencies into nominal subject phrases in English and French, which they argued are backgrounded by default. The authors motivate the backgrounded status of subject phrases using a (corrective) negation test (Erteschik-Shir 1973; Van Valin 1995; Van Valin and LaPolla 1997). The test relies on the intuition that constituents can only be negated or denied if they are contained in the part of the sentence that is asserted/focused. The authors note (p. 19) that '[i]n a neutral context, it is more felicitous to negate (part of) the object than (part of) the subject.' This explains the difference between (6-a) and (6-b).

(6) a. A: The football player liked the color of the car.
B: No, the size of the car.
b. A: The football player liked the color of the car.
B: #No, the baseball player.

As we will see later, it is unclear whether this test reliably diagnoses backgrounded constituents in other constructions, but for the moment we take the distinction at face value. According to Abeillé and colleagues, the relative infelicity of (6-b) indicates that the subject phrase is backgrounded. Therefore, the account predicts that extraction of a *wh*-filler from inside a subject should result in an island effect. No island effects are predicted, however, for RC-dependencies from the same subjects

Across multiple experiments the authors investigated the acceptability of English *wh*- and RC-dependencies with PP fillers (*pied-piping*, as in (7-a) and (7-b)) and NP fillers (*prepositional stranding* as in (7-c) and (7-d)) from definite subject NPs.

(7) a. **Pied-piping from Subject, *Wh*-question**
Of which sportscar did [the color __] delight the baseball player because of its surprising luminance?

b. **Pied-piping from Subject, RC-dependency**
The dealer sold a sportscar, of which [the color __] delighted the baseball player because of its surprising luminance.

c. **P-stranding in Subject, *Wh*-question**
Which sportscar did the [color of __] delight the baseball player because of its surprising luminance?

d. **P-stranding in Subject, RC-dependency**
The dealer sold a sportscar, which [the color of __] delighted the baseball player because of its surprising luminance.

Experiments 2 and 3 of Abeillé et al. (2020) compared sentences such as those above with counterpart sentences in which the *wh*- and RC-fillers were associated with gaps inside NPs in object position (e.g., (8)) and unquestionably ungrammatical baseline sentences (9).[4]

(8) a. **Pied-piping from Object NP, *Wh*-question**
Of which sportscar did the baseball player love [the color __] because of its surprising luminance?

b. **Pied-piping from Object NP, RC-dependency**
The dealer sold the sportscar of which the baseball player loved [the color __] because of its surprising luminance.

c. **P-stranding in Object NP, *Wh*-question**
Which sportscar did the baseball player love [the color of __] because of its surprising luminance?

d. **P-stranding in Object NP, RC-dependency**
The dealer sold the sportscar which the baseball player loved [the color of __] because of its surprising luminance.

(9) a. **Ungrammatical Baseline, *Wh*-question**
*Which sportscar did the baseball player love the color because of its surprising luminance?

b. **Ungrammatical Baseline, RC-dependency**
*The dealer sold a sportscar, which [the color __] the baseball player loved because of its surprising luminance.

The results of the experiments showed that extraction from object phrases was generally more acceptable than from subject phrases, irrespective of dependency type. Differences in the acceptability of extraction from subjects varied by dependency type and by the category of the filler. For *wh*-questions, both pied-piping and P-stranding dependencies were judged as unacceptable as the ungrammatical baseline (9-a). For RC-dependencies, while P-stranding dependencies were judged as unacceptable as the corresponding ungrammatical baseline (9-a), pied-piping dependencies were judged significantly more acceptable and on par with grammatical P-stranding from an object NP (8-b).

Abeillé and colleagues argue that the results broadly support the FBCC. The unacceptability of *wh*-extraction from subject phrases is predicted. The authors also contend that the results of the RC-experiments align with the FBCC. Without any auxiliary assumptions, the FBCC predicts that both pied-piping and P-stranding RC-dependencies into subjects should be acceptable. The prediction for pied-piping is arguably borne out in English (and in French). However, the unacceptability of P-stranding is inconsistent with the simple predictions of the FBCC. To accommodate the P-stranding results, Abeillé and colleagues argue that there is an additional constraint—independent of the FBCC—that renders P-stranding (inside subjects) unacceptable. They speculate that the factor could be grounded in processing difficulty. We find the possible explanations proposed in Abeillé

et al. (2020) unlikely[5], but for the purposes of the paper we remain agnostic as to why there are differences between P-stranding and pied-piping from nominal subjects.

With the caveat above, the acceptability of pied-piped RC-movement from subjects provides suggestive support for the FBCC. As the FBCC is proposed as a general constraint, it is expected to apply beyond subjects to other domains that have been considered islands. The prediction of the FBCC is that—all else equal—any domain that is backgrounded should block *wh*-dependencies, but should permit RC-dependencies. Our experiment tests these general predictions in Norwegian based on three domains: adjuncts, embedded questions, and (existential) RCs. We also test extraction with P-stranding from nominal subjects as an unacceptable baseline against which to compare the results of the other domains.

1.2. Norwegian

Native speakers of Mainland Scandinavian languages such as Norwegian, Swedish, and Danish are consistently reported to accept and produce filler-gap dependencies into domains that were considered islands in many other languages (see, among others, Christensen 1982; Engdahl 1982, 1997; Erteschik-Shir 1973; Lindahl 2017; Maling and Zaenen 1982; Taraldsen 1982). It has been observed that Norwegian permits filler-gap dependencies into embedded questions and (some types of) relative clauses. The following sentences are examples of such dependencies found in a recent corpus study of children's books (Kush et al. 2021, pp. 22, 25):

(10) **Embedded Question**

Han ene typen vet vi jo ikke engang [hva $\_\__i$ heter].
he one guy.DEF know we PRT NEG even what is.called

'That one guy, we don't even know what __ is called.'
≈ 'That one guy, we don't even know the name of.'

(11) **Relative Clause**

Det$_i$ er det ingen$_k$ [som $\_\__k$ vet $\_\__i$]
that is it no.one REL knows

'That, there is no one who knows __.'
≈ 'No one knows *that*.'

The acceptability of sentences such as those above in Norwegian (and Swedish and Danish) has led some researchers to posit parametric differences in *syntactic* islandhood of EQs and RCs in Mainland Scandinavian on the one hand and languages such as English on the other where extraction from EQs and RCs incurs a more reliable cost.[6]

According to these accounts, the underlying structure of EQs and RCs in Mainland Scandinavian makes it possible to move out of EQs and RCs without violating locality rules on movement, thus rendering the data compatible with traditional syntactic accounts (Lindahl 2017; Nyvad et al. 2017; Vikner et al. 2017).

Island-insensitivity beyond EQs and RCs is not as well-established. The formal literature has largely assumed that subjects are islands for all filler-gap dependencies in Norwegian. This assumption has recently received support from experiments that have shown that sentences such as (12) are consistently rated as unacceptable (Bondevik et al. 2021; Kush and Dahl 2020; Kush et al. 2018, 2019).

(12) **Subject**

*Hvilken gutt$_i$ syntes du at [mora til $\_\__i$] var interessant?
which boy thought you that mother.DEF to was interesting

'Which boy did you think the mother of __ was interesting?'

The islandhood of adjuncts is also less often discussed. A reference grammar of Norwegian (Faarlund 1992, p. 117) provides examples of apparently acceptable topicalization out of tensed (temporal) adjunct clauses in (13).[7] However, Bondevik et al. (2021) found that while topicalization from conditional adjuncts did not result in island effects, topicalization from reason and temporal adjunct clauses did. This suggests that a more nuanced understanding of the islandhood of different adjuncts may be required.

(13) **Adjunct**

a. Det$_i$ blir han sint [når jeg sier __$_i$].
that becomes he angry when I say
'That he becomes angry when I say __.'

b. Den saken$_i$ venter vi her [mens de fikser __$_i$].
that case.DEF wait we here while they fix
'That case we wait here while they fix __.'

In sum, prior work shows that filler-gap dependencies are in principle possible into EQs and RCs (and perhaps some adjuncts) in Norwegian.

Though dependencies into EQs, RCs and possibly adjuncts are reported, the acceptability of extraction from different constituents may vary by dependency type (*wh*-movement, relativization and topicalization). The majority of documented examples of extraction from RCs feature topicalization (Taraldsen 1982; see also Engdahl 1997 and Lindahl 2017). In the parsed child-fiction corpus of Norwegian bokmål (part of NorGramBank, see Rosén et al. 2009), Kush et al. (2021) found that all instances of extraction from RCs were topicalization dependencies. Attested examples of extraction from EQs usually feature either RC-movement or topicalization: Kush et al. (2021) found that of the 404 examples of extraction from EQs in their corpus, 319 featured relativization and the remaining 85 examples were topicalization dependencies. *Wh*-question dependencies are conspicuously absent in most collections of naturally occurring examples.[8] The lack of any examples with *wh*-extraction from these domains is potentially surprising given earlier claims that, in principle, nothing blocks such dependencies in Norwegian (e.g., Maling and Zaenen 1982).

Recent judgment studies paint a roughly similar picture: Kush et al. (2018) did not find *wh*-extraction to be acceptable in Norwegian for extraction from subjects, conditional adjuncts, relative clauses, or complex NPs. A smaller island effect was found for *wh*-movement from *whether* EQs. When investigating topicalization on the other hand, Kush et al. (2019) found that contextually-supported topicalization from EQs was acceptable (though topicalization without context did produce an island effect), while judgments of topicalization from RCs were variable. Topicalization from subjects and complex NPs was, however, unacceptable. Interestingly, the authors also found that topicalization from conditional adjuncts did not produce island effects, an effect which Bondevik et al. (2021) replicated. Finally, Kush and Dahl (2020) confirmed that relativization from EQs did not produce island effects.

Given the variation discussed above, we reasoned that Norwegian was a good language in which to systematically test for differences in island effects across dependency type. An added benefit of testing Norwegian is that Norwegian may also offer us the opportunity to isolate discourse-based (or non-structural) factors that influence the acceptability of 'island violations' and that are independent of syntactic constraints in domains such as EQs and RCs, if those domains are assumed to not be syntactic islands.

2. Materials and Methods

2.1. Design and Materials

We ran a preregistered acceptability judgement study that tested Norwegian speakers' intuition about the acceptability of *wh*- and RC-extraction from four syntactic domains: (1) Nominal Subjects; (2) Conditional Adjuncts; (3) Embedded Questions; (4) Existential RCs. The first three domains have been tested in previous experiments, but the current experiment is the first, to our knowledge, to test existential RCs in Norwegian.[9]

The experiment employed the factorial definition of island effects established by Sprouse (2007) and widely used in previous experimental research cross-linguistically (Almeida 2014; Bondevik et al. 2021; Kush et al. 2019; Pañeda et al. 2020; Sprouse et al. 2012, 2016). Test sentences were multiclausal sentences containing a filler-gap dependency. We created test items by manipulating three factors: DISTANCE, STRUCTURE, and DEPENDENCY. DISTANCE had two levels that controlled whether the gap was in the matrix clause or an embedded clause, corresponding to *Short* or *Long* distance between the filler and the gap. STRUCTURE had two levels that controlled whether the embedded clause was or contained an *Island* structure or not (*no Island*). An island effect was defined as the super-additive interaction of DISTANCE and STRUCTURE. DEPENDENCY controlled whether the filler-gap dependency in test sentence was a (*wh-* or RC-dependency).

Our *wh*-dependencies used lexically restricted *wh*-phrases (e.g., *hvilke aktivister* 'which activists') instead of bare *wh*-phrases. RC-dependencies contained the relative pronoun *som* (glossed as REL) and the lexical material of the head matched the filler in corresponding *wh*-dependency sentences.

For RC dependencies we chose to use what we term *demonstrative* RCs such as those in (14). In demonstrative RCs the RC head is definite and is preceded by (i) the pronoun *det* and a tensed version of the verb *være* ('to be'). In such RCs the pronoun *det* can be interpreted as analogous to the demonstrative *that* in the gloss in (14). In such RCs the pronoun/demonstrative is focused, while the head of the RC and and the RC itself are backgrounded. Since the head of the RC is backgrounded, the dependency is suitable for testing the FBCC.

(14) Det var boken$_i$ [som jeg leste $\_\__i$].
It was book.DEF REL I read
'That was the book that I read.'

We chose to use demonstrative RCs in order to avoid introducing extra lexical or semantic material into the matrix clause of RC-dependency sentences that was not in *wh*-dependency sentences. One complication associated with using demonstrative RCs is that they are string-ambiguous with cleft sentences. The sentence in (14) could also be interpreted in the right contexts as roughly analogous to the English *it*-cleft *It was the book that I read*. Clefting in Norwegian places the head of the cleft in focus as in English (Gundel 2002; Hedberg 2000; Prince 1978), so the FBCC predicts that it should not be possible to associate a clefted filler with a gap inside a backgrounded constituent (on par with *wh*-extraction).

We acknowledge that this potential ambiguity potentially complicates using our *wh*- and RC-dependencies to test the divergent predictions of the FBCC for focalizing and non-focalizing dependencies. We note that some of our items give us the opportunity to test whether the ambiguity had negative effects: Our EQ items were adapted from Kush and Dahl (2020), which tested the acceptability of 'eventive' relativization from EQs. Eventive RCs are not subject to the same ambiguity as demonstrative RCs, so to the extent that effects in our study match those in Kush and Dahl (2020), we can conclude that the ambiguity did not cause a problem.

We applied the DISTANCE x STRUCTURE x DEPENDENCY design to all four of the island types mentioned above. We briefly discuss design considerations for each island type in turn.

2.1.1. Subjects

Before testing for cross-dependency differences in extraction from adjuncts, EQs and existential RCs, we wanted to establish an unacceptable baseline against which to compare other effects. Prior work shows that Norwegian speakers consistently rate *wh-* and RC-dependencies from nominal Subjects with P-stranding as unacceptable (Bondevik et al. 2021; Kush and Dahl 2020; Kush et al. 2018 2019). We therefore reasoned that we could use the Subject Island sub-design as an example of uncontroversially unacceptable extraction.

Since we are primarily using the Subject Island items as a benchmark for unacceptability, it is immaterial for the immediate purposes of our study whether the unacceptability arises from a grammatical violation (as is standardly assumed) or whether it reflects parsing difficulties related to P-stranding (as suggested by Abeillé et al. 2020).

Subject Island items were adapted from previous studies, e.g., (Bondevik et al. 2021; Kush and Dahl 2020; Kush et al. 2018). A full example item is presented in (15). Here and in the other items, the *Long-Island* conditions ((15-g) and (15-h)) correspond to sentences where the gap is located inside an island structure.

(15) a. **Short x No Island x *Wh*-dependency**

Hvilke aktivister er redde for at fabrikken skader miljøet?
which activists are worried C factory.DEF harms environment.DEF
'Which activists are worried that the factory is harming the environment?'

b. **Short x No Island x RC-dependency**

Det er aktivistene som er redde for at fabrikken skader miljøet.
those are activists.DEF REL are worried C factory.DEF harms environment.DEF
'Those are the activists that are worried that the factory is harming the environment.'

c. **Long x No Island x *Wh*-dependency**

Hvilken fabrikk er aktivistene redde for at skader miljøet?
which factory are activists.DEF worried C harms environment.DEF
'Which factory are the activists worried __ is harming the environment?'

d. **Long x No Island x RC-dependency**

Det er fabrikken som aktivistene er redde for at skader miljøet.
that is factory.DEF REL activists.DEF are worried C harms environment.DEF
'That is the factory that the activists worry __ is harming the environment.'

e. **Short x Island x *Wh*-dependency**

Hvilke aktivister er redde for at avfall fra fabrikken skader miljøet?
which activists are worried C waste from factory.DEF harms environment.DEF
'Which activists are worried that waste from the factory is harming the environment?'

f. **Short x Island x RC-dependency**

Det er aktivistene som er redde for at avfall fra fabrikken skader
those are activists.DEF that are worried C waste from factory.DEF harms
miljøet.
environment.DEF
'Those are the activists that are worried that waste from the factory is harming the environment.'

g. **Long x Island x *Wh*-dependency**

Hvilken fabrikk er aktivistene redde for at avfall fra skader miljøet?
which factory are activists.DEF worried C waste from harms environment.DEF
'Which factory are the activists worried that waste from __ harms the environment?'

h. **Long x Island x RC-dependency**

Det er fabrikken som aktivistene er redde for at avfall fra skader
that is factory.DEF that activists.DEF are worried C waste from harms
miljøet.
environment.DEF
'That is the factory that the activists are worried that waste from __ is harming the environment.'

2.1.2. Embedded Questions

EQ items were adapted from Kush and Dahl (2020). We used EQs where either *hva* ('what') or *hvor* ('where') were linked to VP-internal gaps. In *Long* test sentences, the gap always occurred in embedded subject position immediately following the complementizer *at* (in the *Long-noIsland* condition) or the *wh*-phrase (in the *Long-Island condition*). Extraction of a subject immediately following a lexically-filled complementizer phrase is acceptable for (most) Norwegians, i.e., Norwegian does not exhibit Comp-t effects (Lohndal 2009; Vangsnes 2019). A full example of a test item is in (16):

(16) a. **Short x No Island x *Wh*-dependency**

Hvilken snekker sa at hylla skulle monteres i stuen?
which carpenter said C shelf.DEF should install.PASS in living.room.DEF
'Which carpenter said that the shelf should be installed in the living room?'

b. **Short x No Island x RC-dependency**

Det var snekkeren som sa at hylla skulle monteres i stuen.
that was carpenter.DEF REL said C shelf.DEF should install.PASS in living.room.DEF
'That was the carpenter that said that the shelf should be installed in the living room.'

c. **Long x No Island x *Wh*-dependency**

Hvilken hylle sa snekkeren at skulle monteres i stuen?
which shelf said carpenter.DEF C should install.PASS in living.room.DEF
'Which shelf did the carpenter say __ should be installed in the living room?'

d. **Long x No Island x RC-dependency**

Det var hylla som snekkeren sa at skulle monteres i stuen.
that was shelf.DEF REL carpenter.DEF said C should install.PASS in living.room.DEF
'That was the shelf that the carpenter said __ should be installed in the living room.'

e. **Short x Island x *Wh*-dependency**

Hvilken snekker sa hvor hylla skulle monteres?
which carpenter said where shelf.DEF should install.PASS
'Which carpenter said where the shelf should be installed?'

f. **Short x Island x RC-dependency**

Det var snekkeren som sa hvor hylla skulle monteres.
that was carpenter.DEF REL said where shelf.DEF should install.PASS
'That was the carpenter that said where the shelf should be installed.'

g. **Long x Island x *Wh*-dependency**

Hvilken hylle sa snekkeren hvor skulle monteres?
which shelf said carpenter.DEF where should install.PASS
'Which shelf did the carpenter say where __ should be installed?'

h. **Long x Island x RC-dependency**

Det var hylla som snekkeren sa hvor skulle monteres.
that was shelf.DEF that carpenter.DEF said where should install.PASS
'That was the shelf that the carpenter said where __ should be installed.'

Our items, such as those from Kush and Dahl (2020), differed from the EQs tested in Kush et al. (2018 2019) in two ways. First, we did not use embedded polar questions (i.e., *whether* questions). Second, our items used the Norwegian equivalents of *know*, *forget*, *say*, *remember*, and *find out* (many of which Lahiri (2002) categorizes as *responsive* predicates)

as embedding predicates instead of *rogative* predicates such as *wonder* which were used in Kush et al. (2018).[10] We chose to use these EQs because Kush et al. (2021) found that dependencies such as (16) were far more frequent in the input than dependencies into polar questions and (ii) Kush and Dahl (2020) found that relativization from such EQs did not result in an island effect. We wished to see whether we would replicate this result.

In order to determine the predictions of the FBCC for EQs, we wanted to establish whether EQs are backgrounded or focused. EQs are traditionally considered backgrounded, insofar as they do not convey the assertion of the clause (Simons 2007). We nevertheless chose to test whether EQs in our items were focused or backgrounded using the negation test employed by Abeillé et al. (2020). We can conclude, for example, that the embedded declarative clause is part of the focus domain in (17-a) because we can negate constituents, such as the subject, in corrective responses (17-b).

(17) a. Snekkeren sa at hylla skulle monteres.
carpenter.DEF said that shelf.DEF should install.PASS
'The carpenter said that the shelf should be installed.'
b. Nei, kommoden.
No dresser.DEF
'No, the dresser.'

Applying the same test to the EQ in (18-a) results in (18-b). We have marked the judgment in (18-b) as '(%)#' to reflect that there is some inter-speaker variation between the Norwegian-speaking authors of the paper and ten additional informants, on whether it is infelicitous to negate the subject in a corrective response. However, seven out ten of our informants reported either complete infelicity for the negation of elements inside *wh*-clauses or noted that negation of the subject in the EQ was less felicitous than negation of the subject in the corresponding embedded declarative clause (17).

(18) a. Snekkeren sa hvor hylla skulle monteres.
carpenter.DEF said where shelf.DEF should install.PASS
'The carpenter said where the shelf should be installed.'
b. %# Nei, kommoden.
No dresser.DEF
'No, the dresser.'

The fact that informants, on balance, judged negation to be less felicitous with the EQ than with an embedded declarative is consistent with there being a difference between the backgroundedness of the two constituents on average. Thus, the FBCC predicts that there should be an observable penalty for extracting a focused *wh*-filler from an EQ compared to RC-extraction from the same EQ. There are two ways to deal with inter-participant variation in the results of the negation test: one could simply ignore it and treat EQs as backgrounded across the board (as a traditional view might assume), or one could assume that (participants' judgments of) the backgroundedness of EQs can vary in a way that should interact with possibility of extraction. Under the first option, the penalty associated with *wh*-extracting from an EQ should be relatively consistent across trials (e.g., it should clearly affect the mode of the judgment distribution). Under the second option, we expect judgments of *wh*-extraction from an EQ to vary across trials or participants, corresponding to whether the EQ is interpreted as backgrounded. We take the first option.

2.1.3. Adjuncts

We used *conditional* clauses headed by *om* 'if' as the adjunct in our items, as in (19-a) below. Adjuncts are traditionally regarded as backgrounded constituents. Again we ran the corrective negation test to determine whether we could confirm the traditional categorization. We asked the same individuals as above whether it was possible to negate the adjunct-internal object *kniven* 'the knife' in the example below, which was based on one of our test items.

(19) a. Kokken blir sur om hun bruker kniven.
chef.DEF gets angry if she uses knife.DEF
'The chef gets angry if she uses the knife.'
b. %# Nei, øsen.
no ladle.DEF
'No, the ladle.'

Once again, we saw some variability in judgments. Overall, seven out of ten informants judged negation in (19-a) to be completely infelicitous our degraded. Following our logic above, we interpret this as suggestive confirmation that conditional adjuncts are backgrounded. As such, the FBCC predicts that there should be an observable penalty for *wh*-extraction from a conditional adjunct compared to RC-extraction.

A full set of items is presented below:

(20) a. **Short x No Island x *Wh*-dependency**

Hvilken kokk misliker at hun bruker den skarpe kniven?
which chef dislikes C she uses the sharp knife.DEF
'Which chef dislikes that she uses the sharp knife?'

b. **Short x No Island x RC-dependency**

Det er kokken som misliker at hun bruker den skarpe kniven.
that is chef.DEF REL dislikes C she uses the sharp knife.DEF
'That is the chef that dislikes that she uses the sharp knife.'

c. **Long x No Island x *Wh*-dependency**

Hvilken kniv misliker kokken at hun bruker?
which knife dislikes chef.DEF C she uses
'Which knife does the chef dislike that she uses __?'

d. **Long x No Island x RC-dependency**

Det er kniven som kokken misliker at hun bruker.
that is knife.DEF REL chef.DEF dislikes C she uses
'That is the knife that the chef dislikes that she uses __.'

e. **Short x Island x *Wh*-dependency**

Hvilken kokk blir sur om hun bruker den skarpe kniven?
which chef gets angry if she uses the sharp knife.DEF
'Which chef gets angry if she uses the sharp knife?'

f. **Short x Island x RC-dependency**

Det er kokken som blir sur om hun bruker den skarpe kniven.
that is chef.DEF REL gets angry if she uses the sharp knife.DEF
'That is the chef that gets angry if she uses the sharp knife.'

g. **Long x Island x *Wh*-dependency**

Hvilken kniv blir kokken sur om hun bruker?
which knife gets chef.DEF angry if she uses
'Which knife does the chef get angry if she uses __?'

h. **Long x Island x RC-dependency**

Det er kniven som kokken blir sur om hun bruker.
that is knife.DEF REL chef.DEF gets angry if she uses
'That is the knife that the chef gets angry if she uses __.'

2.1.4. Relative Clauses

Kush et al. (2018 2019) tested *wh*-extraction and topicalization from RCs that were attached constituents in direct or oblique argument positions such as (21).[11]

(21) Hvilken film$_i$ snakket han med mange kritikere som likte __$_i$?
Which film spoke he with many critics REL liked
'Which film did he speak with many critics that liked __?'

We chose to test extraction from existential RCs such as (22) instead, because existential RCs (alongside clefts) are the RC-type most commonly observed in naturalistic examples of extraction (Engdahl 1997; Erteschik-Shir and Lappin 1979; Kush et al. 2021; Lindahl 2017).

(22) Det var mange som bestilte ølet.
it was many REL ordered beer.DEF
'There were many people who ordered the beer.'

Existential RCs are different from ordinary restrictive RCs in that they introduce or assert the existence of a referent (the head) and use the RC to provide potentially new information about that referent (Engdahl 1997; Lambrecht 1994). [12] Existential RCs are string-ambiguous with cleft sentences in Norwegian, in that both constructions have an expletive subject *det*, followed by the copula. To avoid the possibility that participants interpreted existential RCs as clefts, we used bare (weak) quantifiers as RC-heads (see Milsark 1974), which bias towards an existential reading.

If backgrounded material is that which is not asserted or which is presupposed, then existential RCs are not backgrounded. To verify whether the negation test identifies existential RCs as not-backgrounded, we tested the felicity of negating the RC-internal object *øl/ølet* 'beer/the beer' as in (23-b).

(23) a. Det var mange som bestilte øl/ølet.
it was many REL ordered beer/beer.DEF
'There were many people who ordered (the) beer.'
b. %# Nei, vin/vinen.
No wine/wine.DEF
'No, (the) wine.'

Eight of ten of our informants were willing to accept the negation in (23-b), corroborating the consensus view that existential RCs are not backgrounded. As such, the FBCC predicts that both relativization and *wh*-extraction from RCs in our experiment should be felicitous.

An example item set is below:

(24) a. **Short x No Island x *Wh*-dependency**

Hvilken servitør sa at mange bestilte ølet?
which waiter said C many ordered beer.DEF
'Which waiter said that many people ordered the beer?'

b. **Short x No Island x RC-dependency**

Det var servitøren som sa at mange bestilte ølet.
that was waiter.DEF REL said that many ordered beer.DEF
'That was the waiter that said that many people ordered the beer.'

c. **Long x No Island x *Wh*-dependency**

Hvilket øl sa servitøren at mange bestilte?
which beer said waiter.DEF C many ordered
'Which beer did the waiter say many people ordered __?'

d. **Long x No Island x RC-dependency**

Det var ølet som servitøren sa at mange bestilte.
that was beer.DEF REL waiter.DEF said C many ordered
'That was the beer that the waiter said many people ordered __.'

e. **Short x Island x *Wh*-dependency**

Hvor mange var det som bestilte ølet?
how many was it REL ordered beer.DEF
'How many were there that ordered the beer?'

f. **Short x Island x RC-dependency**

Det var mange som bestilte ølet.
it was many REL ordered beer.DEF
'There were many people that ordered the beer.'

g. **Long x Island x *Wh*-dependency**

Hvilket øl var det mange som bestilte?
which beer was it many REL ordered
'Which beer were there many people that ordered __?'

h. **Long x Island x RC-dependency**

Det var ølet som det var mange som bestilte.
that was beer.DEF REL it was many REL ordered
'That was the beer that there were many people that ordered __.'

Before moving on, we must note one way in which our RC items deviated from the strict factorial design, since it has bearing on whether cross-condition comparisons are apt. A commonality across materials in our Subject, EQ, and Adjunct sub-designs was that *wh*-fillers and RC heads in *Short* conditions were lexical NPs extracted from matrix subject position. This design feature could not be carried over to *Short-Island* conditions in the RC sub-design because the formal subject of an existential RC construction is an expletive *det* that cannot be questioned or relativized. Since it was not possible to have short-distance extraction from subject position in these items, we had to construct alternative comparison sentences. For the *Short Island RC-dependency* condition (24-f), we used the simple existential RC that formed the base used in the other *Island* sentences. In these sentences there simply was no filler-gap dependency in the matrix clause. For the *Short Island Wh-dependency* (24-g), we created a *wh*-question by questioning the quantified head of the base existential RC. Given that the these sentences deviated from the factorial design, the interaction effect that we measure does not offer a direct measurement of a residual island effect where all extraneous factors have been cleanly factored out. We therefore do not rely solely on the presence or absence of a statistically significant interaction to determine whether there was an island effect or not.

2.2. Participants

A total of 96 native Norwegian speakers were recruited through Prolific and public announcements on several social media websites. Prolific participants were paid GBP 3.50; participants recruited via social media were not compensated. The average study duration was 23 min. After completing the experiment, the participants were asked a series of demographic questions that concerned age, their language/dialect background, their parents' language/dialect background, and their preferred standard of written Norwegian. We included a question about participants' age by providing five age groups to choose from.[13] The distribution of participants by age group was the following: 18–30 (54 participants), 31–39 (25 participants), 40–49 (11 participants), 50–59 (2 participants), and 60–69 (4 participants). We excluded 1 participant who reported that Norwegian was not their native language.

2.3. Procedure

A total of 16 items of 8 conditions apiece were created for each island type, according to the design outlined above. This resulted in 512 test sentences that were distributed across 8 experimental lists, with each participant seeing 64 test sentences. The test sentences were interspersed with 40 filler sentences, resulting in 104 sentences that each participant was asked to judge. The filler sentences contained 10 acceptable fillers (*good fillers*) and 30 unacceptable fillers (*bad fillers*) that varied in length and complexity. We chose unequal number of acceptable and unacceptable fillers to compensate for the fact that at least 75% of test items were acceptable sentences without any grammatical errors (*Short-noIsland*, *Short-Island*, *and Long-noIsland* conditions). Adding more unacceptable items allowed us to (roughly) counter-balance the number of acceptable and unacceptable sentences in the whole experiment to mitigate scale bias. Sentences were pseudorandomly ordered between participants, such that no two consecutive items were of the same island or filler type.

The experiment was built using jsPsych (De Leeuw 2015) and hosted on a JATOS server at UiT The Arctic University of Norway (Lange et al. 2015). Participants completed the task using their own personal computer. They were instructed to give ratings to sentences that were presented on a screen one at a time. The judgments were given on a seven point scale. Participants were instructed to treat 1 as *dårlig* 'bad' and 7 as *god* 'good', and to rate sentences that were 'maybe not completely unacceptable, but also not fully acceptable' with a score in the middle range. The first two items of the study were unannounced practice 'filler' items: one regular, acceptable sentence, and one unacceptable sentence. Termed 'anchoring' items by Sprouse and Almeida (2017), these items served to expose participants to, and encourage use of, the entire range of the scale. These items were the same, and presented in the same order, for every participant.

2.4. Analysis

Data preprocessing included three steps. First, ratings were z-transformed by participant to reduce bias from differences in participants' use of the 7-point scale. Second, trials where no rating was recorded (68 trials, constituting 0.7% of all trials) were removed from the dataset.[14] Third, one participant with unusually low ratings to grammatical sentences was removed from the dataset. In the preregistration we planned to remove trials where participants responded in less than 1000 ms, but after removing trials with missing ratings, no trials remained with a reaction time less than this threshold. We had also planned to remove any participant whose mean rating to all trials was less than the midpoint of the 7 pt. scale, but there were no participants who met this criterion.

We applied two different types of models to participant ratings to test for island effects: We applied linear mixed-effects models (LMEMs) to z-scored ratings using the lmerTest package (Kuznetsova et al. 2017) in R (R Core Team 2021). We also analyzed participants' untransformed ratings using cumulative ordinal regression with cumulative link mixed models (CLMMs) implemented using the ordinal package (Christensen 2019). Unlike LMEMs, CLMMs do not assume that numerical judgments are drawn from an ordinal scale and have been argued to be more appropriate for analysis of rating data (Bürkner and Vuorre 2019; Liddell and Kruschke 2018). We present the results of both analyses.

We ran separate models for each island type. All models included DISTANCE, STRUCTURE, and DEPENDENCY and their interactions as fixed effects and a full random effects structure (Barr et al. 2013). If island effects vary by dependency type, we expect a three-way interaction of DISTANCE x STRUCTURE x DEPENDENCY. Centered simple difference coding was used for contrasts: DISTANCE (*Long* = −0.5, *Short* = 0.5); STRUCTURE (*Island* = −0.5, *noIsland* = 0.5); DEPENDENCY (*Wh*-dependency = −0.5; RC-dependency = 0.5). Details of the individual models are provided in Section 3.

We report the size of each interaction effect using a Difference-in-Differences (DD) score (Maxwell and Delaney 2004) calculated on the z-scored ratings. We also perform further (informal) comparisons. First, we compare the average (z-scored) rating of the *Long-Island* conditions to the average ratings of *grammatical fillers* (GF in Figure 1) and the average

ratings of all grammatical items (GI in Figure 1, which included the good fillers, the *Short-noIsland*, *Short-Island*, and *Long-noIsland* conditions), as a way of determining the 'overall' acceptability of individual conditions. Such comparisons are important in light of recent findings that statistically significant island effects have been observed in some languages even when the island-violations are judged to be relatively acceptable (see discussion of so-called 'subliminal island effects' in Almeida (2014); Keshev and Meltzer-Asscher (2019); Pañeda et al. (2020)). Second, we compare the ratings of average z-scores of *Long-Island* conditions within constituent type as a way of assessing whether one dependency type is 'more unacceptable' in the absolute sense than another. Third, we examine the distributions of (z-scored) participant judgments in order to determine whether the average acceptability ratings we observe represent a central tendency in the data and to determine the extent to which there was variability in judgments. Recent work has argued that this kind of distributional analysis helps in drawing inferences about the source of island effects (see Kush et al. 2018, 2019; Pañeda and Kush 2021 for discussion).

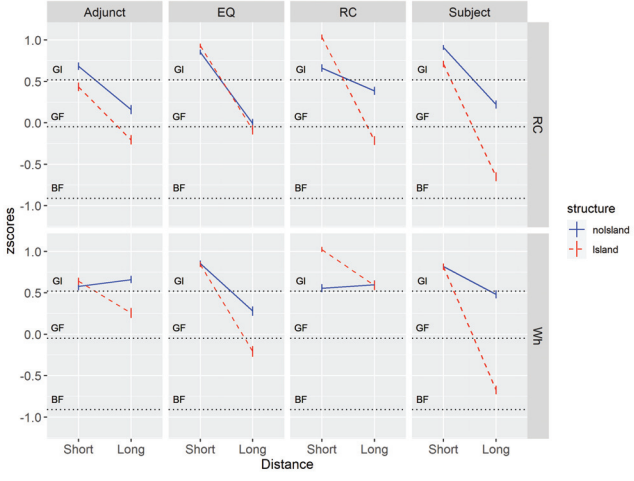


Figure 1. Interaction plots for each island type split by dependency. Error bars represent standard errors. Dotted lines represent mean ratings for all acceptable items ("good" items, GI), acceptable fillers ("good" fillers, GF), and unacceptable fillers ("bad" fillers, BF).

3. Results

Participants rated bad filler sentences low (mean z-score = −0.91). The average rating of bad fillers is marked on each interaction plot with the dotted line labeled 'BF' to give a sense of the lower bound of unacceptability. Good fillers, which varied in complexity, received an average rating close to z = 0, represented by the dotted line labeled 'GF'. Aggregated together all good items (filler and test) were rated close to z = 0.51 ('GI' in Figure 1). Ratings on these trials indicate that the participants understood and performed the task as expected. Below we present the results for each of the island types in turn.

3.1. Subjects

Statistical analysis revealed a significant STRUCTURE x DISTANCE x DEPENDENCY interaction (LMEM: $\beta = 0.52$, $t = 3.27$, $p = 0.0037$; CLMM: $\beta = 2.24$, $z = 3.64$, $p = 0.0003$), indicating that the size of the STRUCTURE x DISTANCE island effect varied across dependency type. Follow-up analysis revealed significant STRUCTURE x DISTANCE interactions for RC-dependencies (LMEM: $\beta = -0.67$, $t = -5.66$, $p < 0.0001$; CLMM: $\beta = -1.19$, $z = -2.61$, $p = 0.0090$) and *Wh*-dependencies (LMEM: $\beta = -1.18$, $t = -10.4$, $p < 0.0001$; CLMM: $\beta = -3.24$, $z = -6.65$, $p < 0.0001$). The STRUCTURE x DISTANCE interaction effect was larger for *wh*-dependencies than for RC-dependencies (DD = 1.15 v. DD = 0.67, respec-

tively). The difference in size of the interaction effect appears largely driven by the reduced average acceptability of the RC-dependency in the *Long-noIsland* condition (z = 0.22) compared to the *wh*-dependency (z = 0.48). The average acceptability of *wh*-movement from a subject (z = −0.66) and RC-movement from a subject (−0.65) did not differ significantly.

Ratings distributions by condition are presented in Figure 2. Ratings across *Short* conditions were nearly all at the top end of the scale (z ∼ 1). Ratings were differently distributed in the *Long-noIsland* versus *Long-Island* conditions. Ratings in *Long-noIsland* conditions were largely distributed around z = 1, though there was a longer left tail indicating that participants rated the occasional *Long-noIsland* sentences as degraded. In contrast, the *Long-Island* conditions mostly grouped around the lower end of the scale (z < −1), indicating that participants overwhelmingly perceived the sentences as deeply unacceptable.

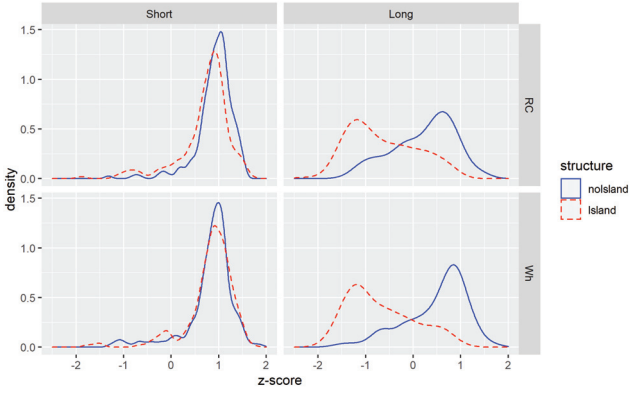


Figure 2. Distributions of ratings for sentences in the Subject Island sub-design split by dependency type and condition.

3.2. Embedded Questions

Statistical analysis revealed a significant STRUCTURE x DISTANCE x DEPENDENCY interaction in the LMEM (β = 0.34, t = 2.05, p = 0.0508), but the 3-way interaction was only marginally significant in the CLMM (β = 1.22, z = 1.82, p = 0.0682). Resolving the three-way interaction revealed that while there was an island effect for *wh*-dependencies as manifested by a significant STRUCTURE x DISTANCE interaction (LMEM: β = −0.48, t = −4.33, p = 0.0003; CLMM: β = −1.37, z = −3.59, p = 0.0003), no such effect was found for RC-dependencies. The DD score was larger for *wh*-dependencies than for RC-dependencies (DD = 0.47 v. DD = 0.15, respectively).

Visual inspection of Figure 1 suggests that the difference in interaction size across dependency type is mostly due to differences in the acceptability of the *Long-noIsland* conditions (*Wh*: z = 0.28 v. RC: z = −0.01), not differences between the *Long-Island* conditions. The average acceptability of *wh*-movement from an EQ (z = −0.21) is relatively close to the mean acceptability of RC-movement (z = −0.08) and post hoc comparisons revealed that the numerical difference between the conditions was not significant ($p > 0.1$).

Ratings distributions are presented in Figure 3. Ratings in *Short* conditions were nearly all high, whereas ratings in *Long* conditions were more variable. The variable ratings of RC-dependencies in the *Long-Island* and *Long-noIsland* conditions overlap completely, confirming that participants did not perceive RC-movement from EQs as marked compared to RC-movement from embedded declaratives. For *wh*-dependencies, ratings in the *Long* conditions were also variable, but there was slightly less overlap between the *Long-Island* and *Long-noIsland* distributions. On the one hand, participants were slightly less likely to give high ratings to *wh*-extraction from EQs than *wh*-extraction from embedded declaratives. This could be interpreted as evidence for a penalty. On the other hand, if we compare

judgments of *Long-Island* sentences across dependency type, we see that the distributions in the *Long-Island Wh* and the *Long-Island RC* condition are nearly identically distributed. This could be taken to suggest that participants did not perceive *wh*-movement from EQs to be worse, in the absolute sense, than RC-movement from EQs.

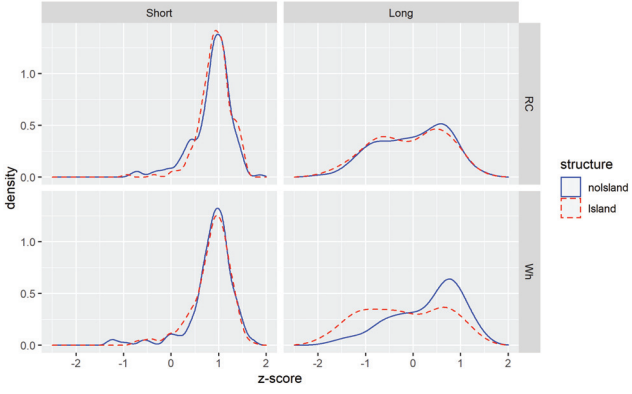


Figure 3. Distributions of ratings for sentences in the Embedded Question sub-design split by dependency type and condition.

3.3. Adjuncts

Statistical analysis revealed a significant STRUCTURE x DISTANCE x DEPENDENCY interaction in the CLMM ($\beta = 1.36$, $z = 2.33$, $p = 0.0199$), though the effect was only marginally significant in the LMEM ($\beta = 0.32$, $t = 1.84$, $p = 0.0819$). Effect sizes differed between the two dependency types (DD = 0.47 for *wh*-dependencies v. DD = 0.11 for RCs). We ran a separate analysis for each dependency type, which revealed an absence of an island effect for RC-dependencies as manifested by a non-significant STRUCTURE x DISTANCE interaction (LMEM: $p = 0.5$; CLMM $p = 0.9$). Visual inspection of Figure 1 confirms the absence of an island effect for RC-dependencies. There was a significant STRUCTURE x DISTANCE interaction for *wh*-dependencies (LMEM: $\beta = -0.41$, $t = -2.61$, $p = 0.0214$; CLMM: $\beta = -1.35$, $z = -3.11$, $p = 0.0019$). The interaction is notable, however, in that *wh*-movement from a conditional adjunct was rated higher on average ($\approx$ z = 0.25) than RC-movement ($\approx$ z = -0.21). Post-hoc comparisons revealed that this difference was significant ($p < 0.05$).

Figure 4 shows that the participants' ratings across *Short* conditions were generally rated high ($\sim$+1). The distribution of judgments in *Long* conditions differed across dependency type. Judgments in the the *Long-noIsland-Wh-dependency* condition were mostly high, similar to judgments in *Short* conditions. Judgments in the the *Long-Island-Wh dependency* condition were more variable. The distribution suggests relatively polar responses across trials with a larger cluster around z = +0.75 and a smaller cluster around z = -1. It seems that the majority of trials were rated around +0.75, suggesting that the sentences were judged acceptable more often than they were rejected. Ratings of *Long* sentences for RC-dependencies had qualitatively different distributions. Ratings of *Long-noIsland* sentences had a mode at the top of the scale, but many sentences were rated as less acceptable to some degree. In the corresponding *Long-Island-RC dependency* condition, the ratings are centered around the midpoint of the scale with substantial variance. If we compare judgments in the *Long-Island* conditions across dependency type, it appears that participants were more likely to give a high acceptability score to *wh*-movement from a conditional than RC-movement, despite the fact that an 'island effect' is only observed with *wh*-movement.

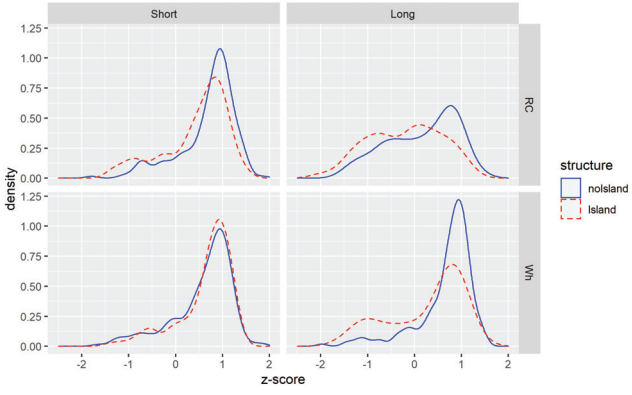


Figure 4. Distributions of ratings for sentences in the Adjunct Island sub-design split by dependency type and condition.

3.4. Relative Clauses

For Relative Clauses, the STRUCTURE x DISTANCE x DEPENDENCY interaction was significant in the LMEM ($\beta = -0.44$, $t = -2.29$, $p = 0.0368$) and marginally significant in the CLMM ($\beta = -2.16$, $z = -1.89$, $p = 0.0587$). Resolving the three-way interaction revealed STRUCTURE x DISTANCE interactions for both RC- (LMEM: $\beta = -0.95$, $t = -9.22$, $p < 0.0001$; CLMM: $\beta = -5.48$, $z = -5.3$, $p < 0.0001$) and *wh*-dependencies (LMEM: $\beta = -0.51$, $t = -2.8$, $p = 0.0145$; CLMM: $\beta = -2.95$, $z = -3.44$, $p = 0.0006$). The interaction observed for RC-dependencies resembles a standard island effect, such that the *Long-Island* condition is rated significantly worse than the *Long-noIsland* condition. The interaction with *wh*-dependencies does not resemble the typical interaction pattern. First, there is not a signficant difference between the average acceptability of the *Long-Island* and *Long-noIsland* conditions. The interaction appears to be driven entirely by extremely high acceptability ratings in the *Short-Island* condition. We attribute the high ratings to the relative simplicity of the structures used in these conditions. As discussed in Section 2.1, we were forced to deviate from a strict factorial design in the *Short-Island* condition. Therefore it seems inappropriate to use DD scores to quantify the 'RC island effect'. Instead, the most informative comparison for determining whether there is an island effect is to compare the mean ratings in the *Long-Island* (z = 0.59) and *Long-noIsland* (z = 0.60) conditions. We interpret the negligible difference between the two *Long* conditions as evidence that there is no island effect for *wh*-extraction from an existential RC.

Rating distributions by condition are presented in Figure 5. Similar to other domains, *Short* conditions received consistently high ratings. Looking at *wh*-dependencies where we observed no island effect, we see that the *Long-Island* and *Long-noIsland* distributions are nearly identical, indicating that participants did not distinguish *wh*-movement from a declarative complement clause from an existential RC. Interpreting the ratings of *Long* RC-dependencies is less straightforward. Participants generally rated sentences from the *Long-noIsland* condition high, indicating that they judged RC-movement from a declarative complement clause acceptable. Ratings of RC-movement from existential RCs, however, show considerable variation and no clear mode. Insofar as the distribution is clearly different from the *Long-noIsland* condition, the conclusion that there is an island effect of some sort is supported. It seems, however, that the island effect does not reflect uniform rejection of the dependencies (as seen with movement from subjects).

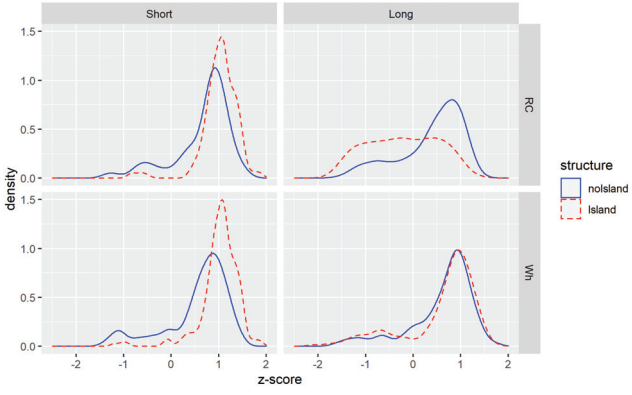


Figure 5. Distributions of ratings for sentences in the Relative Clause sub-design split by dependency type and condition.

4. Discussion

We found that statistically significant DISTANCE x STRUCTURE effects varied by domain and dependency type. These 'island' effects indicate that some extractions resulted in decreases in acceptability that could not be accounted for by main effects of STRUCTURE and DISTANCE alone. We found that significant effects (i) often reflect highly variable judgments in the 'island-violating' *Long-Island* condition and (ii) do not always entail that 'island violations' are unacceptable in absolute terms. In what follows we discuss effects by domain and how our results align with predictions of the FBCC.

4.1. Subjects

We observed large island effects for both RC- and *wh*-extraction from the subject phrases we tested. We saw that the size of the island effects differed by dependency type, but we reasoned that the statistically significant differences were not practically or theoretically meaningful in that participants reliably rejected RC- and *wh*-dependencies into subjects. Regardless of its origins, the subject island effect provides a benchmark for a large, consistent island effect against which we can compare other effects in the study.

We do not draw conclusions about whether the island effect we observed is consistent with the predictions of the FBCC because our items used preposition stranding, which Abeillé et al. (2020) argued was unacceptable for independent reasons. We point out, however, that if preposition stranding causes the problem, the explanation for the unacceptability cannot be that readers could not locate the gap site. The stranded preposition marked the gap site very clearly. It is also unlikely that the explanation can be linked to a preference for pied-piping, since pied-piping is not an option in Norwegian RC-dependencies, and it is not used in *wh*-questions in standard varieties.

4.2. Embedded Questions

Replicating the findings of Kush and Dahl (2020), we found that relativization of a subject from an EQ did not result in a significant island effect. We observed a significant island effect for *wh*-extraction from the same EQs, though this island effect was smaller (DD = 0.49) than our subject island effects (DDs = 1.14). Since we replicate the absence of an island effect for relativization, we conclude that the ambiguity between demonstrative relativization and clefting did not have an effect on the acceptability of extraction from EQs.

Although there was an island effect for *wh*-movement, the effect was largely due to differences in the average acceptability of *wh*- and RC-extraction from *declarative* complements. The average acceptability of *wh*-extraction from EQs was not significantly different from

RC-extraction from the same EQs. Further, judgments of *wh*-extraction and relativization from EQs exhibited nearly identical variability.

If EQs are more backgrounded than declarative complement clauses, the FBCC predicts that we should see a penalty for *wh*-extraction from an EQ compared to *wh*-extraction from a declarative complement. A comparable penalty should not be observed for RC-extraction. A proponent of the FBCC might interpret the island effect we observed as consistent with this prediction.

We think, however, that there are also reasons to treat the interaction with caution. First, the small interaction effect could simply be an artifact of a ceiling effect. As discussed, the interaction emerges for *wh*-dependencies because there is a pairwise difference between the *Long-noIsland* and the *Long-Island* conditions, but not between the *Short* conditions. However, both *Short* conditions are rated essentially at the top of the scale, where potentially meaningful acceptability differences may be compressed. Second, the the average acceptability ratings and their distributions in the *Long-Island* condition were nearly identical for *wh*- and RC-extraction. The similarities make it hard to conclude that participants perceived *wh*-extraction as 'worse' than RC-extraction.

4.3. Adjuncts

We found that relativization from a conditional adjunct did not result in a significant island effect, similar to English results from Sprouse et al. (2016).

Wh-extraction yielded a statistically significant island effect, but the effect was small (DD = 0.44) because the mean rating of *wh*-extraction from a conditional ($z \approx 0.25$) was relatively high. It was above the average rating of the good fillers in the experiment and significantly higher than the average rating of relativization from a conditional adjunct. Thus, *wh*-extraction from conditionals appears to be, on average, 'acceptable' despite the island effect. The distribution of judgments confirmed that most participants considered *wh*-extraction from an adjunct to be acceptable more often than not: Participants rated the sentences near the top of the scale on the majority of trials, though they rated the sentences at the bottom end of the scale on the rest of trials.

We now turn to how our results square with the FBCC. The absence of an island effect for relativization from conditional adjuncts is consistent with the FBCC insofar as the FBCC does not predict island effects for relativization from *any* constituent. The significant island effect for *wh*-extraction is potentially consistent with the FBCC.

Once again, we think that the interaction effect, and the judgment distributions underlying that effect, do not unequivocally support the FBCC. We saw that the relatively high mean rating of *wh*-extraction from an adjunct was the result of averaging over a judgment distribution that had a mode at the top of the scale and a smaller proportion or judgments at or below zero. That is, participants were more likely, on balance, to judge *wh*-extraction from an adjunct just as acceptable as from an embedded declarative. If conditional adjuncts are uniformly backgrounded, we would expect a reliable penalty for *wh*-extraction from them: participants should have rated *wh*-extraction from an adjunct to be less acceptable than from a declarative on a majority of trials. This is not what we see. It seems instead that insofar as there is a penalty, it is observed inconsistently, on a small number of trials.

A proponent of the FBCC could accommodate the inconsistent unacceptability of *wh*-extraction, by letting the backgroundedness of conditional adjuncts vary. Under this interpretation, participants rated *wh*-extraction from conditional adjuncts acceptable on trials where they interpreted the conditional as part of the focus domain and rejected *wh*-extraction on trials where they interpreted the adjunct as backgrounded. If variability in backgroundedness is behind the judgment variability we observed, there is a simple prediction: there should be a negative correlation between individual items' backgroundedness as measured by the negation test and the acceptability of *wh*-movement from those adjuncts.[15] We have not conducted the experiments to confirm or falsify this prediction, but have made our items and data publicly available on the project's OSF page to any researchers who are interested in conducting the experiments.

Finally, it should be noted that our results, which seem to suggest that *wh*-extraction from a conditional is largely acceptable, appear to conflict with the results of Kush et al. (2018), where *wh*-extraction from conditional adjuncts resulted in large, consistent island effects across three experiments. What is responsible for the differences in extractability? We do not have an iron-clad explanation for the discrepancy, but we suspect that lexical differences between items used in the studies may have played a role: The current experiment adapted adjunct items from Bondevik et al. (2021), which differed from those used in Kush et al. (2018) in two potentially relevant ways. First, items in Bondevik et al. (2021) were constructed relative to a context sentence, which may have indirectly led to more 'natural-sounding' items than those used in Kush et al. (2018). Second, items in Bondevik et al. (2021) and our study used a very restrictive set of predicates in the main clause. In all *Island* conditions, the matrix verb was *bli* ('become'), followed by an adjective describing an emotional state (e.g, 'happy', 'angry', 'nervous' and 'surprised'). In Kush et al. (2018) a wider set of matrix predicates was used ('complain', 'sigh', 'protest', 'worry' and 'become happy'). If the matrix predicate influences the possibility of extraction from an adjunct, as suggested by Truswell (2011) and others, the difference in predicate types could be the source of the apparent discrepancy in results. We encourage more systematic investigation of how different predicates influence the possibility of extracting from conditionals and other adjuncts and whether the observed cross-dependency differences in English would be attenuated with different predicates.

4.4. Relative Clauses

Participants rated *wh*-extraction from an existential RC just as acceptable as *wh*-extraction from a declarative complement clause. However, they rated relativization from an existential RC as significantly worse, on average, than relativization from a declarative complement. Where judgments of *wh*-extraction were consistently acceptable, judgments of relativization exhibited a large degree of variation, ranging across the scale from $z = -1$ to $z = +1$.

As we discussed in the *Materials* section, existential RCs are non-presuppositional and are therefore not backgrounded. As such, the FBCC predicts that they should therefore allow *wh*-extraction. Our results are consistent with this prediction.

The island effect for RC-movement from existential RCs does not follow from any formalized account that we are aware of. According to the FBCC, RC-movement should, all else equal, be permissible wherever *wh*-movement is possible. Therefore, the source of the island effect must lie elsewhere. We do not have a concrete proposal for what additional factor(s) could be at play, but our results rule out a simple explanation grounded in complexity or dependency length. One possibility is that it is specifically the combination of demonstrative relativization and an existential RC that causes infelicity or unacceptability. If so, we might predict that sentences with eventive relativization would not be judged as unacceptable:

(25) Jeg likte faktisk ølet$_i$ som det var mange som hata __$_i$
I liked actually beer.DEF REL it was many REL hated
lit. 'I actually liked the beer that there were many who hated __.'
≈ 'I actually liked the beer that many hated.'

The variation in judgments also suggests that RC-movement from existential RCs may not be uniformly unacceptable. It is possible that item-specific factors, individual differences, or some interaction of the two modulate acceptability. For example, participants may have struggled (to varying degrees) to accommodate/imagine a supporting context for relativization across individual items (see Chaves and Putnam 2020 for more discussion). Providing a formal foundation for these intuitions should be one goal of future inquiry.

5. Conclusions

Our results show that *wh*- and RC-dependencies into nominal subjects are consistently unacceptable in Norwegian, but judgments of extraction from other domains show more nuanced patterns. We observed small island effects for *wh*-extraction from conditional adjuncts and embedded questions, but not for relativization from the same constituents. We argued, however, that the mere presence of significant island effects for *wh*-movement did not straightforwardly support the FOCUS BACKGROUND CONFLICT constraint. Our results also suggest that other semantic/pragmatic factors above and beyond a simple focus-background partitioning are needed to explain cross-dependency differences in the acceptability of extraction (from domains such as existential RCs). We hope that the data we have collected can be used in the development of more fine-grained accounts of the factors that influence the acceptability of filler-gap dependencies in 'island' environments.

Author Contributions: Conceptualization, D.K., T.L., A.K., C.S.; methodology, D.K., C.S., T.L., A.K. and P.T.R.; formal analysis, A.K., P.T.R., M.V. and D.K.; data curation, M.V.; writing—original draft preparation, A.K., C.S., P.T.R., M.V., T.L. and D.K.; writing—review and editing, A.K., D.K, C.S. P.T.R., M.V. and T.L.; visualization, A.K.; supervision, D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by an Onsager Fellowship from NTNU to D.K.

Institutional Review Board Statement: Ethical review and approval were waived for this study in accordance with the policies of the Norwegian Centre for Research Data (NSD) because the study collected no personal data which could be used to identify individual participants (either directly or by combining data).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: https://osf.io/ma9jp/.

Acknowledgments: We are grateful to three anonymous reviewers for their constructive and helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BCI	Backgrounded Constituents are Islands
C	complementizer
DEF	definite
EQ	embedded question
FBCC	The Focus-Background Conflict Constraint
NP	noun phrase
PASS	passive
PP	prepositional phrase
RC	relative clause
REL	relative pronoun

Notes

[1] We use the common 'filler-gap' terminology and indicate the 'gap' site of extracted elements for ease of exposition. Such description does not necessarily entail commitment to an analysis where the relation between the filler and the verb *read* in (1-a) and (1-b) is mediated via an empty element such as a trace or where the filler is related directly to the head itself as it would be in trace-less theories (e.g., HPSG, Pollard and Sag 1994); LFG, (Kaplan and Bresnan 1995), or Construction Grammar (Goldberg 1995).

[2] An alternative is that the constraints apply to the structure-building mechanisms themselves. See Nunes and Uriagereka (2000); Stepanov (2007); Uriagereka (1999).

[3] According to Abeillé et al. (2020), the size of the acceptability decrease can vary—presumably corresponding to the *degree of infelicity*.

[4] Other baseline sentences were used in these experiments, but we omit them from discussion to focus on the critical comparisons.

[5] For example, we do not think that there is *less* ambiguity about the true gap site for an extracted PP than for an extracted DP/NP associated with a stranded preposition. If anything, the stranded preposition provides a clearer signal for the true gap site. As McInnerney and Sugimoto (2022) note, there is no independent motivation to assume that P-stranding should ever be harder than pied-piping, since the former is clearly the preferred option for extraction in English. We refer the interested reader to McInnerney and Sugimoto (2022) for an interesting alternate explanation of Abeillé and colleagues' results, according to which the apparent cases of pied-piping from subjects are not cases of extraction at all, but rather instances of base-generated topic PPs.

[6] It has been observed that sometimes extraction from relative clauses such as those in (11) is significantly less degraded in languages outside of the Mainland Scandinavian group including Hebrew, Italian, and even English (see, e.g., Cinque 2010; Kluender 1998; Kush et al. 2013; Lindahl 2017; Rubovitz-Mann 2000; Sichel 2018; Vincent 2021). However, recent experimental work suggests that the perceived acceptability of such sentences does not reflect a complete amelioration of island effects (see, e.g., Vincent 2021).

[7] We have converted Faarlund's examples into the Bokmål written standard to align with other examples across the paper.

[8] A recent corpus study by Abeillé and Winckel (2020) found similar asymmetry for extraction out of subjects in French.

[9] The preregistration can be found at https://osf.io/zksh6/.

[10] A reviewer notes that the complements of verbs such as *know* and *say* can be interpreted as resolved questions, whereas the complements of rogative verbs such as *wonder* are interpreted as 'open' questions in the terminology of Ginzburg and Sag (2000). Insofar as the 'resolved' questions are interpreted as factive complements, they are potentially predicted to behave as more backgrounded than 'open' questions. Contrary to this prediction, island effects are more consistently observed with rogative verbs than with responsive verbs (Abrusán 2014; Pañeda and Kush 2021; Suñer 1991; Torrego 1984).

[11] Christensen and Nyvad (2014) tested similar sentences in Danish.

[12] In essence, existential RCs are similar in semantic content to mono-clausal declaratives such as '*Many ordered beer*'. Interestingly, Norwegian uses existential and presentational RCs (as well as clefts) more often than simple declaratives, presumably reflecting a preference to keep indefinites or new material out of surface subject position (see Diesing 1992; Johansson 2001).

[13] We did not collect exact ages to minimize the amount of potentially-identifiable data we collected.

[14] Due to a software error, participants were not prevented from proceeding to the next trial before giving a response. The small number of affected trials were relatively evenly distributed across participants, items, conditions, and place in the study.

[15] Ambridge and Goldberg (2008) argue that ease of extraction from complement clauses correlates negatively with degree of backgroundedness as measured by the negation test, but in a recent larger study Liu et al. (2022) failed to replicate this finding.

References

Abeillé, Anne, and Elodie Winckel. 2020. French subject island? empirical studies of dont and de qui. *Journal of French Language Studies* 30: 275–300. [CrossRef]

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204: 104293. [CrossRef] [PubMed]

Abrusán, Márta. 2014. *Weak Island Semantics*. Oxford: Oxford University Press.

Almeida, Diogo. 2014. Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *Revista da ABRALIN* 13: 55–93. [CrossRef]

Ambridge, Ben, and Adele Goldberg. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19: 357–89. [CrossRef]

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [CrossRef] [PubMed]

Bondevik, Ingrid, Dave Kush, and Terje Lohndal. 2021. Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics* 44: 223–54. [CrossRef]

Bürkner, Paul-Christian, and Matti Vuorre. 2019, Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science* 2: 77–101. [CrossRef]

Chaves, Rui P., and Michael T. Putnam. 2020. *Unbounded Dependency Constructions*. Oxford: Oxford University Press. [CrossRef]

Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*. Edited by Morris. Halle, Stephen R. Anderson and Paul Kiparsky. New York: Holt, Rinehart and Winston, pp. 232–86.

Chomsky, Noam. 1977. On wh-movement. In *Formal Syntax*. Edited by Peter W. Culicover, Thomas Wasow and Adrian Akmajian. New York: Academic Press, pp. 71–132.

Chomsky, Noam. 1986. *Barriers*. Cambridge: MIT Press.

Chomsky, Noam. 2000. Minimalist Inquiries: The Framework. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Edited by Roger Martin, David Michaels and Juan Uriagereka. Cambridge: MIT Press, pp. 89–155.

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2014. On the nature of escapable relative islands. *Nordic Journal of Linguistics* 37: 29–45. [CrossRef]

Christensen, Kirsti Koch. 1982. On multiple filler-gap constructions in Norwegian. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabet Engdahl and E. Ejerhed. Stockholm: Almquist & Wiksell, pp. 77–98.
Christensen, Rune H. B. 2019. Ordinal—Regression Models for Ordinal Data. Available online: https://CRAN.R-project.org/package=ordinal (accessed on November 26).
Cinque, Guglielmo. 1990. *Types of A-Bar Dependencies*. Cambridge: MIT Press.
Cinque, Guglielmo. 2010. On a selective 'violation' of the complex NP constraint. In *Structure Preserved: Studies in Syntax for Jan Koster*. Edited by Jan Wouter Zwart and Mark de Vries. Amsterdam: John Benjamins, pp. 81–90.
De Leeuw, Joshua R. 2015. jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods* 47: 1–12. [CrossRef] [PubMed]
Diesing, Molly. 1992. *Indefinites*. Cambridge: MIT Press.
Engdahl, Elisabet. 1982. Restrictions on unbounded dependencies in Swedish. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabet Engdahl and Eva Ejerhed. Stockholm: Almquist & Wiksell, pp. 151–74.
Engdahl, Elisabet. 1997. Relative clause extractions in context. *Working Papers in Scandinavian Syntax* 60: 51–79.
Erteschik-Shir, Nomi. 1973. On the Nature of Island Constraints. Ph.D. dissertation, MIT, Cambridge, MA, USA.
Erteschik-Shir, Nomi, and Shalom Lappin. 1979. Dominance and the functional explanation of island phenomena. *Theoretical Linguistics* 6: 41–86. [CrossRef]
Faarlund, Jan Terje. 1992. *Norsk syntaks i funksjonelt perspektiv*. Oslo: Universitetsforlaget.
Ginzburg, Jonathan, and Ivan A. Sag. 2000. *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. Stanford: CSLI Publications.
Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
Gundel, Jeanette K. 2002. Information structure and the use of cleft sentences in English and Norwegian. In *Information Structure in a Cross-Linguistic Perspective*. Edited by Hilde Hasselgård, Stig Johansson, Bergljot Behrens and Cathrine Fabricius-Hansen. Amsterdam: Brill Rodopi, pp. 113–28. [CrossRef]
Hedberg, Nancy. 2000. The referential status of clefts. *Language* 76: 891–920. [CrossRef]
Huang, C. T. James. 1982. Logical Relations in Chinese and the Theory of Grammar. Ph.D. dissertation, MIT, Cambridge, MA, USA.
Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.
Johansson, Mats. 2001. Clefts in contrast: A contrastive study of it clefts and wh clefts in English and Swedish texts and translations. *Linguistics* 39: 547–82. [CrossRef]
Kaplan, Ronald M., and Joan Bresnan. 1995. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In *Formal Issues in Lexical-Functional Grammar*. Edited by Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III and Annie Zaenen. Stanford: Stanford University, pp. 1–102.
Keshev, Mayaan, and Aya Meltzer-Asscher. 2019. A processing-based account of subliminal wh-island effects. *Natural Language and Linguistic Theory* 37: 621–57. [CrossRef]
Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In *The Limits of Syntax*. Edited by Peter W. Culicover and Louise McNally. Leiden: Brill, pp. 241–79.
Kuno, Susumu. 1987. *Functional Syntax: Anaphora, Discourse and Empathy*. Chicago: University of Chicago Press.
Kush, Dave, Akira Omaki, and Norbert Hornstein. 2013. Microvariation in islands? In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 239–64.
Kush, Dave, and Anne Dahl. 2022. L2 transfer of L1 island-insensitivity: The case of Norwegian. *Second Language Research* 38: 315–346. [CrossRef]
Kush, Dave, Charlotte Sant, and Sunniva Briså Strætkvern. 2021. Learning island-insensitivity from the input: A corpus analysis of child- and youth-directed text in Norwegian. *Glossa: A Journal of General Linguistics* 6: 1–50. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language & Linguistic Theory* 36: 743–779. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95: 393–420. [CrossRef]
Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82: 1–26. [CrossRef]
Lahiri, Utpal. 2002. *Questions and Answers in Embedded Contexts*. Oxford: Oxford University Press.
Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
Lange, Kristian, Simone Kühn, and Elisa Filevich. 2015. "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS ONE* 10: e0130834. [CrossRef]
Liddell, Torrin M., and John K. Kruschke. 2018. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology* 79: 328–48. [CrossRef]
Lindahl, Filippa. 2017. Extraction from Relative Clauses in Swedish. Ph.D. dissertation, University of Gothenburg, Gothenburg, Sweden.
Liu, Yingtong, Rachel Ryskin, Richard Futrell, and Edward Gibson. 2022. On dissociating adjunct island and subject island effects. *Cognition* 222: 104902. [CrossRef] [PubMed]

Lohndal, Terje. 2009. Comp-t effects: Variation in the position and features of c. *Studia Linguistica* 63: 204–32. [CrossRef]
Maling, Joan, and Annie Zaenen. 1982. A phrase structure account of Scandinavian extraction phenomena. In *The Nature of Syntactic Representation*. Edited by Pauline Jacobson and Geoffrey K. Pullum. Dordrecht: Springer Netherlands, pp. 229–82. [CrossRef]
Maxwell, Scott E., and Harold D. Delaney. 2004. *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, 2nd ed. New Jersey: Lawrence Erlbaum.
McInnerney, Andrew, and Yushi Sugimoto. 2022. On dissociating adjunct island and subject island effects. *Proceedings of the Linguistic Society of America* 7: 5207. [CrossRef]
Milsark, Gary. 1974. Existential Sentences in English. Ph.D. dissertation, MIT, Cambridge, MA, USA.
Nunes, Jairo, and Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 13: 20–43. [CrossRef]
Nyvad, Anne Mette, Ken Ramshøj Christensen, and Sten Vikner. 2017. Cp-recursion in Danish: A cP/CP-analysis. *The Linguistic Review* 34: 449–77. [CrossRef]
Pañeda, Claudia, and Dave Kush. 2022. Spanish embedded question islands effects revisited: An experimental study. *Linguistics* 60: 463–504. [CrossRef]
Pañeda, Claudia, Sol Lago, Elena Vares, João Veríssimo, and Claudia Felser. 2020. Island effects in Spanish comprehension. *Glossa: A Journal of General Linguistics* 5: 21. [CrossRef]
Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
Prince, Ellen F. 1978. A comparison of wh-clefts and it-clefts in discourse. *Language* 54: 883–906. [CrossRef]
R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Rosén, Victoria, Paul Meurer, and Koenraad De Smedt. 2009. LFG Parsebanker: A toolkit for building and searching a treebank as a parsed corpus. Paper presented at the Seventh International Workshop on Treebanks and Linguistic Theories, Groningen, The Netherlands, January 23–24. pp. 127–33.
Ross, John Robert. 1967. Constraints on Variables in Syntax. Ph.D. dissertation, MIT, Cambridge, MA, USA.
Rubovitz-Mann, Talia. 2000. Extractions from Relative Clauses: An Information-Structural Account. Ph.D. dissertation, Hebrew University of Jerusalem, Jerusalem, Israel.
Schütze, Carson T., Jon Sprouse, and Ivano Caponigro. 2015. Challenges for a theory of islands: A broader perspective on Ambridge, Pine, and Lieven. *Language* 91: e31–e39. [CrossRef]
Sichel, Ivy. 2018. Anatomy of a counterexample: Extraction from relative clauses. *Linguistic Inquiry* 49: 335–78. [CrossRef]
Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117: 1034–56. [CrossRef]
Sprouse, Jon. 2007. A Program for Experimental Syntax: Finding the Relationship between Acceptability and Grammatical Knowlege. Ph.D. dissertation, University of Maryland, College Park, MD, USA.
Sprouse, Jon, and Diogo Almeida. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: A Journal of General Linguistics* 2: 14. [CrossRef]
Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34: 307–44. [CrossRef]
Sprouse, Jon, Matt Wagers, and Colin Phillips. 2012. A test of the relation between working memory and syntactic island effects. *Language* 88: 82–124. [CrossRef]
Stepanov, Arthur. 2007. The end of CED? minimalism and extraction domains. *Syntax* 10: 80–126. [CrossRef]
Suñer, Margarita. 1991. Indirect questions and the structure of CP: Some consequences. In *Current Studies in Spanish Linguistics*. Edited by Héctor Campos and Fernando Martínez-Gil. Washington, DC: Georgetown University Press, pp. 283–312.
Taraldsen, Knut Tarald. 1982. Extraction from relative clauses in Norwegian. In *Readings on Unbounded Dependencies in Scandinavian Languages*. Edited by Elisabet Engdahl and Eva Ejerhed. Stockholm: Almquist & Wiksell, pp. 205–21.
Torrego, Esther. 1984. On inversion in Spanish andsome of its effects. *Linguistic Iinquiry* 15: 103–29.
Truswell, Robert. 2011. *Events, Phrases, and Questions*. Oxford: Oxford University Press.
Uriagereka, Juan. 1999. Multiple spell-out. In *Working Minimalism*. Edited by Samuel David Epstein and Norbert Hornstein. Cambridge: MIT Press, pp. 251–82.
Van Valin, Robert. 1995. Towards a functionalist account of so-called 'extraction constraints'. In *Complex Structures: A Functionalist Perspective*. Edited by Betty Divriendt, Louis Goossens and Johan van der Auwera. Berlin: Mouton de Gruyter, pp. 29–60.
Van Valin, Robert D., and Randy J. LaPolla. 1997. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.
Vangsnes, Øystein A. 2019. Comp trace effects across North Germanic varieties. In *The Sign of the V—Papers in Honour of Sten Vikner*. Edited by Ken Ramshøj. Christensen, Henrik Jørgensen, and Johanna L. Wood. Aarhus: Department of English, School of Communication & Culture, Aarhus University, pp. 657–83. [CrossRef]
Vikner, Sten, Ken Ramshøj Christensen, and Anne Mette Nyvad. 2017. V2 and cP/CP. In *Order and Structure in Syntax I: Word Order and Syntactic Structure*. Edited by Laura R. Bailey and Michelle Sheehan. Berlin: Language Science Press, pp. 313–24.
Vincent, Jake Wayne. 2021. Extraction from Relative Clauses: An Experimental Investigation into Variable Island Effects iin English—or—This Is a Dissertation that We Really Needed to Find Someone Who'd Write. Ph.D. dissertation, University of California Santa Cruz, Santa Cruz, CA, USA.

languages

MDPI

Article

The Radical Unacceptability Hypothesis: Accounting for Unacceptability without Universal Constraints

Peter W. Culicover [1,2,*], Giuseppe Varaschin [3] and Susanne Winkler [4]

[1] Department of Linguistics, The Ohio State University, Columbus, OH 43210, USA
[2] Department of Linguistics, University of Washington, Seattle, WA 98195, USA
[3] Institut für deutsche Sprache und Linguistik, Humboldt University of Berlin, 10099 Berlin, Germany; giuseppe.varaschin@gmail.com
[4] Englishes Seminar, University of Tübingen, 72074 Tübingen, Germany; susanne.winkler@t-online.de
* Correspondence: culicover.1@osu.edu

Abstract: The Radical Unacceptability Hypothesis (RUH) has been proposed as a way of explaining the unacceptability of extraction from islands and frozen structures. This hypothesis explicitly assumes a distinction between unacceptability due to violations of local well-formedness conditions—conditions on constituency, constituent order, and morphological form—and unacceptability due to extra-grammatical factors. We explore the RUH with respect to classical islands, and extend it to a broader range of phenomena, including freezing, A′ chain interactions, zero-relative clauses, topic islands, weak crossover, extraction from subjects and parasitic gaps, and sensitivity to information structure. The picture that emerges is consistent with the RUH, and suggests more generally that the unacceptability of extraction from otherwise well-formed configurations reflects non-syntactic factors, not principles of grammar.

Keywords: syntactic theory; island constraints; processing complexity; unacceptability and grammaticality; A′ constructions; frequency; surprisal

Citation: Culicover, Peter W., Giuseppe Varaschin, and Susanne Winkler. 2022. The Radical Unacceptability Hypothesis: Accounting for Unacceptability without Universal Constraints. *Languages* 7: 96. https://doi.org/10.3390/languages7020096

Academic Editors: Anne Mette Nyvad, Ken Ramshøj Christensen, Juana M. Liceras and Raquel Fernández Fuertes

Received: 29 November 2021
Accepted: 6 April 2022
Published: 13 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Syntactic islands are syntactic configurations that in principle should permit extraction, but appear not to. A typical example is (1), which illustrates the unacceptability of extracting from a relative clause.

(1) a. Sandy read [$_{NP}$ a book [$_{S}$ that deals with economic theory]].
b. * What subject$_i$ did Sandy read [$_{NP}$ a book [$_{S}$ that deals with t_i]]

It is characteristic of islands that they *appear* to be well-formed, in that all LOCAL CONSTRAINTS ON FORM are satisfied. For example, in (1b) the wh-phrase *what subject* is in clause-initial position, where it should be in a *wh*-question. There is a gap in the complement position of the preposition that determines its function and allows the subcategorization requirements of the preposition to be met. All of the phrases are otherwise well-formed: e.g., the various categories are in the correct linear order and all conditions on subcategorization and morphological agreement are satisfied.

In the absence of a plausible alternative, linguists have hypothesized that the unacceptability of (1b) reflects a violation of a syntactic constraint on extraction from a relative clause configuration. Unlike the constraints that determine linear order, subcategorization and agreement, this constraint is NON-LOCAL in nature because the gap can be embedded at an arbitrary depth within the relative clause, as (2b) illustrates:

(2) a. * What subject$_i$ did Sandy read [$_{NP}$ a book [$_{S}$ that reveals [$_{S}$ that Kim worked on t_i]]]?

b. * What subject$_i$ did Sandy read [$_{NP}$ a book [$_S$ that reveals [$_S$ that Taylor knows … [that Kim worked on t_i]]]]?

Any syntactic account of phenomena like (2) will typically require grammars of natural languages to include constraints whose domain of application goes well beyond local trees or phrases, encompassing pieces of structure that, though finite in principle, have no upper-bound (Kaplan and Zaenen 1995; Pullum 2019). A corollary of this is that the description language one uses to state syntactic constraints must be endowed with special devices that accomplish the feat of finitely characterizing the unbounded disjunction of paths that may separate a filler from its corresponding gap (devices like existential quantification over nodes or variables in the sense of early transformational grammar).

Ross (1967) showed that these constraints on extraction were general, and not features of particular rules or constructions. Given their abstract nature, a reasonable hypothesis is that such constraints are universal properties of the language faculty, and govern all constructions involving extraction. This hypothesis has driven much of syntactic theorizing since Ross (1967) and the option of attributing the unacceptability that results from violating constraints on extractions to general grammatical principles remains active in much contemporary theorizing (Bošković 2015; Chomsky 2001, 2008; Citko 2014; Nunes and Uriagereka 2000; Phillips 2013a, 2013b; Rizzi 1990; Sabel 2002; Villata et al. 2016, i.a.).[1]

However, a plausible case can be made that these constraints are simply descriptive generalizations. On this view, certain syntactic configurations give rise to unacceptability without violating conditions on grammatical form (Boeckx 2008, p. 154). In fact, at this point there is a substantial literature that makes the case that many constraints on extraction do not reflect violation of grammatical principles, but non-syntactic factors such as processing complexity (Arnon et al. 2005; Chaves 2013, 2020; Chaves and Dery 2014, 2019; Chaves and Putnam 2020; Culicover 2013b, 2013c; Deane 1991; Goldberg 2006; Hofmeister et al. 2007, 2013a; Hofmeister and Sag 2010; Hofmeister et al. 2013b; Kluender 1991, 1992, 1998, 2004; Kluender and Kutas 1993b; Newmeyer 2016; Sag et al. 2006, 2007; Staum Casasanto et al. 2010, i.a.).

In this article we pursue this idea, extending the Radical Unacceptability Hypothesis of Culicover and Winkler (2018, p. 380):

Radical Unacceptability Hypothesis (RUH): ll judgments of reduced acceptability in cases of otherwise well-formed (i.e., locally well-formed) extractions are due to processing complexity, not syntactic constraints.

The basic idea is that processing complexity is responsible for a broader class of judgments of unacceptability beyond islands per se. Processing complexity arises from such factors as parsing A′ chains, referential processing and the management of information structure. We focus specifically on acceptability judgments which result from A′ extractions (*wh*-movement, topicalization, etc.) from 'strong' islands and other configurations from which A′ extractions are allegedly *never* allowed, such as relative clauses and subjects. The phenomena that we cite here are primarily those that we have addressed in our own prior work, in many cases complementing other research in the field.

This article is organized as follows. First we sketch out in Section 2 a picture of the relationship between acceptability judgments, on the one hand, and the various factors that determine these judgments. We take the position that unacceptability neither directly nor necessarily reflects ungrammaticality, in the sense of a violation of a grammatical condition. From this perspective, an understanding of the ways in which acceptability judgments may arise is essential in investigating the nature of grammar.

In Section 3 we discuss the theoretical basis for the distinction between grammaticality and acceptability. We also briefly review the classical island constraints of Ross (1967), pointing to the substantial literature that shows that these constraints are at best descriptive generalizations about phenomena that are better explained in terms of non-syntactic factors.

In Sections 4 and 5 we review patterns of unacceptability that do not all fall under the classical island constraints and argue that these, likewise, are not explained in terms of grammatical constraints, but non-syntactic factors. Among the phenomena that we consider are: freezing (Section 4.1), A′ chain interactions (Section 4.2), topic islands (Section 4.3), zero relative clauses (Section 4.4), weak crossover (Section 5.1), parasitic gaps (Section 5.2), and sensitivity to information structure (Section 5.3).

Section 6 addresses phenomena for which accounts in terms of the RUH are prima facie incompatible with the RUH; we suggest ways in which they may ultimately be brought under the RUH.

Finally, on the basis of our review of the causes of unacceptability in cases of extraction, we conclude in Section 7 that there is strong evidence for the following extended version of the RUH.[2]

Extended Radical Unacceptability Hypothesis (ERUH): All judgments of reduced acceptability in cases of otherwise well-formed (i.e., locally well-formed) extractions are due to non-syntactic factors, not grammatical constraints.

2. Sources of Unacceptability

Let us consider the reasons for a judgment that a sentence is less than fully acceptable. Clearly, violation of a grammatical condition is one source of such a judgment. For example, in (3a) the verb and its complement are in the wrong order, in (3b) there is a subcategorization problem, while in (3c) there is a failure of subject-verb agreement.

(3) a. * Sandy the beer drank;
b. * Sandy relies about Kim;
c. * Sandy are happy.

Such linear order, subcategorization, and morphological agreement constraints are what we call LOCAL WELL-FORMEDNESS CONDITIONS (LWFC). A LWFC, as we understand it, is a constraint on a local piece of linguistic structure, such as adjacent sister nodes or mother-daughter configurations in a tree of depth-1. What defines a LWFC is the fact that it applies to structures of a pre-determined maximum finite size; within some frameworks, these may extend beyond local trees to include non-recursive clausal structures or sequences of phrasal projections, e.g., X′ structures, understood as trees of depth-3 (Jackendoff 1977).

How does violation of an LWFC produce a judgment of unacceptability? The obvious answer is that the form of the example is incompatible with the form stipulated by the LWFC. It is useful to think of LWFCs in terms of experience and expectations. Speakers' prior exposure to their language contributes to the emergence of probabilistic expectations regarding what structures they are likely to hear next. Some of these expectations become consistent and stable enough so they can be described in terms of symbolic LWFCs (Bybee 2006, 2010; Bybee and Hopper 2001; Culicover 2005, 2015; Culicover and Nowak 2003). A LWFC is established on the basis of experience with examples that share certain characteristics, for example, that the order of a VP in English is V > NP, not NP > V. If a given example has these characteristics, then its form is expected on the basis of experience. But if it does not have these characteristics, then its form is surprising, and this leads to the judgment of unacceptability.

We assume, therefore, that there is a relationship between the degree of surprise triggered by a linguistic form, or SURPRISAL, and acceptability. Low surprisal corresponds to high levels of acceptability, higher levels of surprisal correspond to lower levels of acceptability (Hale 2001, 2003; Levy 2005, 2008, 2013; Levy and Jaeger 2007; Park et al. 2021). Surprisal is inversely related to frequency: the higher the frequency of a construction in a given context, the lower its surprisal; the lower the frequency of a construction in a context, the higher its surprisal.[3]

Clearly, the frequency of experience plays a role in determining the level of surprisal even when productive LWFCs are not at stake. There are special cases in English where the order NP > V is possible in VP, e.g., (4).

(4) One swallow does not a summer make.

This example contrasts sharply with (3a). For speakers who accept it, it is because they have encountered it in their experience; it is a special construction in their grammar (Culicover 2021). This experience leads to the probability of hearing the verb *make* follow the NP object *a summer* being much higher than it is for NP > V sequences in general. As a result, surprisal in the case of (4) is lower than it is in the case of the structurally identical (3a), and acceptability is higher.

So we have the relationship shown in Figure 1. Experience increases the frequency of particular constructions, and lack of experience corresponds to zero frequency. Frequency leads to expectations. Some of the expected patterns can be described as general LWFCs (i.e., principles of grammar), and some cannot, as we discuss below. Regardless of this, conformity to expectations leads to low surprisal, and low surprisal corresponds to acceptability.[4]

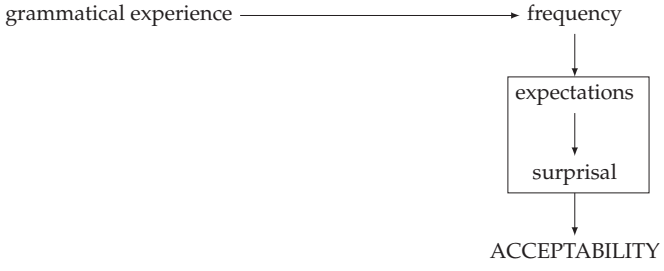


Figure 1. The logic of acceptability judgments for grammatical conditions.

Having established this relationship between grammatical experience and judgments of acceptability, we can now consider other sources of acceptability judgments. One source can be found in the early literature in generative grammar, which suggested that some instances of unacceptability may result from processing complexity and not grammar (e.g., Chomsky 1965; Jackendoff and Culicover 1972; Miller and Chomsky 1963). In particular Miller and Chomsky (1963) demonstrated clearly that unacceptability can arise due to processing complexity in a sentence that satisfies all LWFCs, arguably due to limitations of short-term memory.

It is plausible to assume that higher complexity leads to lower frequency, hence greater surprisal. Since LWFCs can themselves be understood as emergent byproducts of experience-driven expectations, we anticipate that high complexity should have a similar effect on judgments as violation of LWFCs. We therefore extend our picture to that in Figure 2.

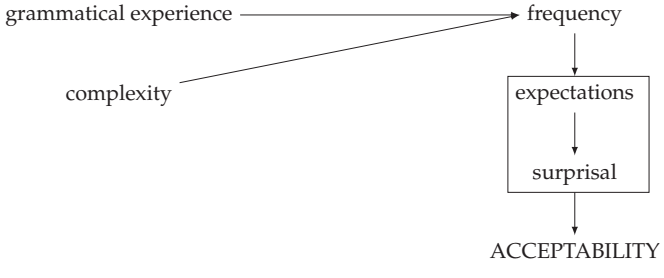


Figure 2. The logic of acceptability judgments for grammatical conditions, version 2.

As we proceed, we flesh out 'complexity' with a number of more specific factors.

Given this general framework, it is now possible to understand a wide range of cases of unacceptability judgments as responses to surprisal. Where the expectations come from that lead to such judgments is a complex question, and each case has to be evaluated on its own terms. In the discussion to follow we offer some suggestions, as well as pointers to relevant literature, recognizing that we are far from understanding all of the fine details. The property that is common to all sources of unacceptability is that lack of conformity to expectations leads to surprisal. In other words, surprisal acts like a CAUSAL BOTTLENECK between a wide range of independent factors that impinge on speakers' expectations and a (behaviorally measurable) acceptability response (Levy 2008).

3. The Acceptability/Grammaticality Distinction and Standard Island Constraints

We suggested above that classical islands of the kind discovered by Ross (1967) may simply be useful generalizations about the kinds of extraction patterns that yield a high level of surprisal, giving rise to an unacceptability response from speakers. If in fact these island patterns are simply generalizations, the following question arises: what factors lead to such generalizations? One answer to this question is the RUH, which in the present framework amounts to the claim that the surprisal associated with island violations stems from the influence of non-syntactic factors in the frequency of particular structures. This hypothesis explicitly assumes a distinction between unacceptability due to violations of local well-formedness conditions (LWFCs)—conditions on constituency, linear order and morphological form—and unacceptability due to non-syntactic factors such as processing complexity as outlined in Section 2.

This distinction has a long lineage in the history of generative grammar (see, for example, Bever 1970; Chomsky 1965; Jackendoff and Culicover 1972; Miller and Chomsky 1963 for some early instances). As soon as language came to be viewed as a cognitive capacity integrated within the larger ecology of the mind, linguists were quick to speculate that grammatical constraints are not the only factors that contribute to the acceptability of sentences (e.g., Kluender 1991, 1998; Kluender and Kutas 1993b). Acceptability came to be viewed as a psychological effect that could be triggered by a host of disparate factors, grammaticality being just one among them (Chomsky 1965, pp. 11–12).

The first exploration of this idea was Miller and Chomsky's (1963) account of the unacceptability of multiple center-embedding structures (e.g., *the man who the boy who the students recognized pointed out is a friend of mine* (Chomsky 1965) in terms of short-term memory limitations. The first attempt to apply this rationale to constraints on extraction was Jackendoff and Culicover's (1972) proposal to explain the restrictions to movement out of ditransitive VPs in terms of perceptual strategies for identifying A′ dependency gaps. Their basic idea was that structures like (5a) are unacceptable because the verb-adjacent NP superficially satisfies the verb's selectional requirement, and the parser expects a gap after the preposition *to* as in (5b)—this is arguably a type of 'garden path'; see Pritchett (1988) for a range of examples. In terms of the model summarized in Figure 3, the absence of a preposition after the NP in (5a) contradicts the frequency-based expectations of the speaker, and, therefore, yields a surprisal effect that contributes to unacceptability.

(5) a. * Who$_i$ did Taylor give t_i a book?
b. Who$_i$ did Taylor give a book to t_i?

In order to explain these phenomena in purely grammatical terms, it would be necessary to enrich the language for stating syntactic constraints in non-trivial ways.[5] Rather than appealing to ad hoc extensions, non-syntactic accounts along the lines of work cited above in Section 1 promise to allow us to keep syntactic theory reasonably simple and constrained. Given their potential to make syntax simpler, it is only natural that we consider the possibility that in some cases the unacceptability of extraction from classical islands reflects not grammar, but processing complexity that arises from particular syntactic configurations, as the RUH proposes.

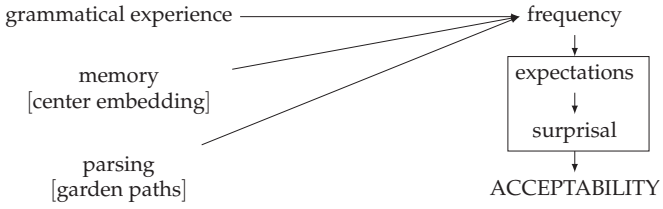


Figure 3. The logic of acceptability judgments for grammatical conditions, version 3.

The application of RUH to classical islands is inspired by two general observations. First, classical island constraints are, in general, TOO STRONG: they exclude sentences that are actually judged to be acceptable by speakers in many circumstances.[6] As an illustration, consider the Complex NP Constraint discussed in connection to (1) above. The counterexamples to this constraint provided below come from Erteschik-Shir and Lappin (1979, p. 58), Pollard and Sag (1994, p. 206) and Sag (1997, p. 454).

(6) a. This is the kind of weather$_i$ that there are [$_{NP}$ many people [$_S$ who like t_i]].
b. Which diamond ring$_i$ did you say that there was [$_{NP}$ nobody in the world [$_S$ who could buy t_i]]?
c. There were several old rock songs$_i$ that she and I were [$_{NP}$ the only ones [$_S$ who knew t_i]]?

Second, classical island constraints are also TOO WEAK: they fail to exclude extraction patterns that speakers generally consider to be unacceptable. In Sections 4 and 5 we review several examples of A′ extractions that do not fall under the classical accounts of islands but which, nonetheless, are unacceptable (Chomsky 1973, 1977, 1986, 2008; Ross 1967, i.a.).

Furthermore, most, if not all, island constraints appear to function in a wide range of languages, and may be universal. If so, the question arises as to the source of such universals. Evolution is an unlikely explanation; island constraints are neither undecomposable features of language that could have arisen by a simple random mutation streamlined by economy constraints (like Merge Labeling and Agree are claimed to be (Berwick and Chomsky 2016; Chomsky et al. 2019)), nor the kinds of features that could have been selected for by adaptive pressures, leading to a gradual evolutionary process (Corballis 2017; Jackendoff 1999; Pinker and Bloom 1990; Progovac 2016). It is, therefore, implausible that the human linguistic phenotype evolved specifically to exclude extraction from all of the specific configurations that have been proposed as islands in the literature. One alternative is that the causes of unacceptability in extractions are what biologists call SPANDRELS: phenotypic traits that are not directly selected, but emerge as byproducts of a complex interaction of independent functional adaptations (Gould and Lewontin 1979). In the case of islands, these may be general cognitive factors related to memory (Kluender and Kutas 1993b), attention (Deane 1991), and the management of information flow in discourse (Erteschik-Shir 1977, 2007; Erteschik-Shir and Lappin 1979). [7]

Chaves and Putnam (2020) offer an extended discussion of classical islands. They review substantial evidence that virtually all of these allow acceptable violations. In addition, they document the factors that enter into judgments of unacceptability (see also Newmeyer 2016). The case they make supports the RUH as an alternative to the default syntactic approach to unacceptability of islands.[8] To further support this view, in the next sections we review briefly a number of additional phenomena that fall outside of the traditional island constraints, or that are not traditionally categorized as islands, and argue that they too reflect non-syntactic factors. The conclusion that we draw is an extension of the RUH – if the sentence containing an extraction is locally well-formed and unacceptable, the unacceptability must be due to a non-syntactic factor.

4. Processing A′ Chains

In this section, we will explore how several extra-grammatical factors related to the processing and parsing of A′ chains increase processing complexity. This, in turn, contributes to reducing the frequency of the particular A′ configurations in which these factors are manifested. According to the model outlined in Figure 3, lower frequency leads to higher surprisal and reduced acceptability.

4.1. Freezing

Classic freezing, noted first by Ross (1967, p. 305) is exemplified by the relative unacceptability of extracting from an extraposed prepositional phrase, as in (7b).

(7) a. You saw [a picture t_j] yesterday [$_{PP}$ of Thomas Jefferson]$_j$.
b. * Who$_i$ did you see [a picture t_j] yesterday [$_{PP}$ of t_i]$_j$?

Historically, explanations for freezing focus on identifying properties of the syntactic configurations from which extraction is not possible and a corresponding grammatical constraint that explicitly blocks such extraction (Corver 2017). For example, Ross (1967) formulated the Frozen Structure Constraint in (8).

(8) a. The Frozen Structure Constraint: If a clause has been extraposed from a noun phrase whose head noun is lexical, this noun phrase may not be moved, nor may any element of the clause be moved out of that clause. (Ross 1967, p. 295)[9]
b. If a prepositional phrase has been extraposed out of a noun phrase, neither that noun phrase nor any element of the extraposed prepositional phrase can be moved. (Ross 1967, p. 303)

Later, Wexler and Culicover (1980) proposed the Raising Principle and the Freezing Principle, based on considerations of language learnability. The Freezing Principle has the effect of blocking extraction from an extraposed PP, as in (7). The Raising Principle blocks extraction from a constituent raised from a lower clause, as in (9).

(9) a. * Who$_i$ did you say that [friends of t_i]$_j$, you dislike t_j? (subextraction from embedded topicalization)
b. * Who$_i$ did you say that [friends of t_i]$_j$ t_j dislike you? (subextraction from subject)

In (9a) a constituent is extracted from a topicalized constituent. Attribution of the unacceptability in (9b) to the Raising Principle of course depends on an analysis in which the subject is taken to be raised from its clause.[10]

The main point about constraints such as these is that they are categorical. In contrast, Hofmeister et al. (2015) and Culicover and Winkler (2018) argue on the basis of experimental evidence that the unacceptability of so-called 'freezing' configurations is gradient and reflects processing complexity, determined by such factors as DEPENDENCY LENGTH of filler-gap chains and the INTERACTION of overlapping A′ chains.

Regarding the first, in the string *read the book*, there is a minimal dependency between *the* and *book*, and a slightly longer dependency between *read* and *book*. Work such as Gibson (1998, 2000) has suggested that longer dependency distance correlates with processing complexity. As far as we know, there is no consensus on how to measure dependency length; several measures of dependency length have been proposed in the literature, including as a function of number of intervening words (Gibson 1998; Lewis and Vasishth 2005; Liu 2008; Liu et al. 2017; Temperley 2007), of complexity of branching structure (Hawkins 1994, 2004, 2014), and of number of new discourse referents (Gibson 2000). Research has shown that in general languages tend to minimize the distance between dependent elements, measured in terms of hierarchical structure (Futrell et al. 2015; Hawkins 1994, 2004, 2014; Liu 2008; O'Grady et al. 2003; Yadav et al. 2021).[11]

Dependency length is added in Figure 4.

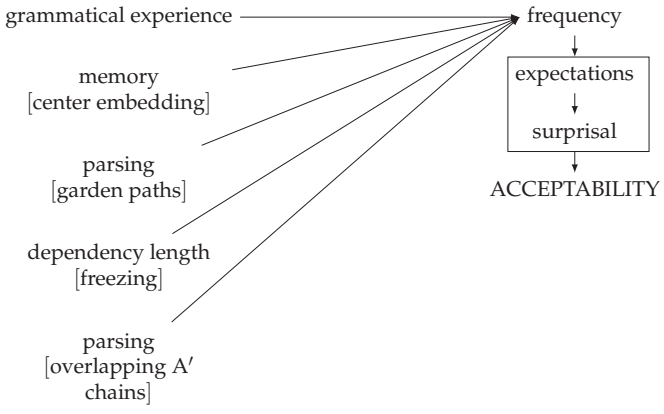
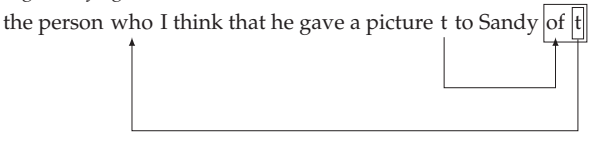


Figure 4. The logic of acceptability judgments for grammatical conditions, version 4.

4.2. Overlapping A′ Chains

Regarding chain interaction, note that in the case of (7), for example, the configuration is that of 'right surfing' (10), where the tail of the extraposed constituent precedes the tail of the chain of the extracted *wh*-phrase in the linear order.[12]

(10) *Right Surfing*



Hofmeister et al. (2015) and Culicover and Winkler (2018) provide experimental evidence that the unacceptability of extraction from extraposed PP depends on the length of the A′ chain and the extraposition chain. The acceptability of the A′ chain alone is a linear function of the length of the dependency, as is the acceptability of PP extraposition alone. The acceptability of extraction from extraposition is determined by the sum of the two overlapping dependencies. Therefore there is no reason to believe that the most unacceptable cases are ungrammatical in a strict sense, to be ruled out by a syntactic constraint. Following the early insights of Miller and Chomsky (1963), the reasoning here presupposes that syntactic constraints as such are largely insensitive to quantitative properties of structures, such as the SIZE of a phrase, the NUMBER of embeddings or the LENGTH of a chain. If acceptability is sensitive to these factors, this is prima facie evidence that the source of the judgment is non-syntactic—plausibly related to working memory capacity.

Similar results were found for the freezing of Heavy NP Shift by Konietzko et al. (2018), as illustrated in (11). This is another case of RIGHT SURFING, where the trace of the constituent that appears in VP-final position contains the trace of the A′ constituent.

(11) a. You put [a picture of FDR]$_j$ on the table.
b. You put t_j on the table [a picture of FDR]$_j$.
c. * Who$_i$ did you put t_j on the table [a picture of t_i]$_j$?

The experimental results reported in Konietzko et al. (2018) suggest, again, that the unacceptability of extraction from the heavy NP is a function of the interaction of the overlapping chain dependencies, and not the configuration of the VP.

To the extent that multiple dependencies entail complexity, the model in Figure 3 leads us to expect that structures with multiple interacting chains will be progressively less frequent in a way that is inversely related to the total size of the interacting chains they contain. As a result, such structures are associated with high surprisal and, therefore, are expected to give rise to low acceptability. We summarize these results by adding the factor 'parsing' to Figure 4.[13]

Why multiple dependencies affect processing complexity is very much an active research question. The most explicit computational models that we are aware of that go beyond the formulation of constraints on parsers are those that appeal to interactions between activation and retrieval from memory, attentional focus, and activation decay (Lewis 1993, 1996; Lewis and Vasishth 2005; Lewis et al. 2006; van Dyke and Lewis 2003; Vasishth and Lewis 2006; Vasishth et al. 2019). No doubt a more fine-grained understanding of the processes involved in the computation of chain dependencies will shed considerably more light on the various phenomena that we have noted here.

4.3. Topic Islands

Topic island phenomena (Rochemont 1989) arguably reflect the interaction of chains in processing as well. Classical examples are given in (12).

(12) a. * What$_i$ does John think that Bill$_j$, Mary gave t_i to t_j?
b. * This is the man who$_i$ that book$_j$, Mary gave t_j to t_i.
c. * How$_i$ did you say [that the car$_j$, Bill fixed t_j t_i]?
d. * This book$_i$, I know that Tom$_j$, Mary gave t_i to t_j.
(Rochemont 1989, p. 147)

Rochemont's account of the unacceptability of examples such as these relies on Chomsky's (1973) Subjacency condition, which blocks movement from a too deeply embedded position in the structure. Depth of embedding is determined by counting the number of barriers, where the notion 'barrier' is defined in terms of a variety of government called L-marking (Chomsky 1986).

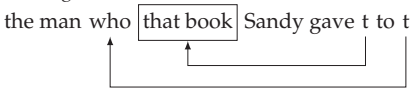
An experimental study by Jäger (2018) confirms that extraction from embedded clauses in which topicalization has occurred is unacceptable. However, Jäger also demonstrates that topicalization alone is less acceptable than canonical SVO order in embedded clauses. Thus, it is plausible that the lower acceptability of embedded topicalization added to the processing cost of long A′ extraction is sufficient to account for the unacceptability of examples like (12).

It is noteworthy that the examples in (12) involve overlapping chain interactions. What is topicalized in the embedded clause is an argument, and requires a trace in its canonical position. If we modify these examples as in (13) so that what appears in initial position in the embedded clause is a sentential adjunct (shown with underlining), acceptability increases. Crucially, a sentential adjunct can be interpreted as soon as it is encountered and does not have to form a chain.

(13) a. ? What$_i$ does John think [that at the concert, Mary proposed to sing t_i]?
b. ? This is the man [who$_i$ at the party, Mary insulted t_j].
c. ? How$_i$ did you say [that when he came home, Bill was feeling t_i]?
d. ? This book$_i$, I know [that if the Times recommends it, Mary will buy t_i].

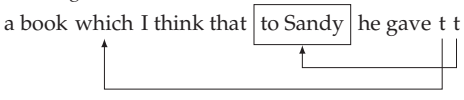
The chain interactions in (12) are different from those seen in the case of freezing. The latter are instances of right surfing, while the former are NESTING, illustrated in (14). In nesting, the fronted constituents are in reverse order to the traces that they form chains with, as shown in (14).

(14) *Nesting*

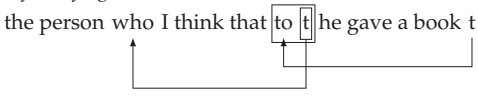


Like right surfing, nesting requires overlapping processing of two chains. Multiple chain processing is also required for CROSSING, illustrated in (15), and the LEFT SURFING configuration, illustrated in (16). In the more acceptable cases of crossing, the fronted constituents are in the same order as the traces that they form chains with, while in left surfing a constituent is extracted from a left extracted constituent.

(15) *Crossing*



(16) *Left Surfing*



Reasoning from the analogy of the freezing experiments, we expect that the processing of multiple overlapping chains to be more difficult than the processing of a single chain or of non-overlapping chains, and correspondingly more unacceptable. We expect the unacceptability to reflect the length of the overlapping chains. As suggested for nesting and crossing, the arrangement of the A′ constituents with respect to their chains is also likely to play a role. Additional complications may arise when a preposition is stranded internally to another constituent, as in the case of left surfing illustrated in (16).[14]

To our knowledge, these factors have not been investigated systematically in the literature. Lewis (1993) proposed a computational model to account for the effects of multiple chains on processing complexity, but his model has not been further developed or brought to bear on the full range of chain interactions discussed here. While it is premature to rule out the possibility that there are grammatical constraints that account for the unacceptability of left surfing, crossing, and nesting, a processing explanation is promising and deserves a focused effort. For a review of recent proposals, see Chaves and Putnam (2020).

Another type of complexity associated with chain interactions is the extent to which the structure that the sentence processor assigns to a string faithfully reflects its semantic structure. This degree of CONGRUENCE determines how easily it is mapped to a semantic interpretation (Culicover and Nowak 2002). In part the ease of this mapping is determined by the extent to which constituents that are adjacent in the string correspond to semantic objects that form a larger semantic object. For example, an adjacent verb and NP in the string are more easily processed as a transitive predicate than a verb and a displaced NP. More complexity in processing would arise if parts of the NP were distributed to non-adjacent positions before and after the verb.[15]

Figure 5 reflects the contribution of congruence to complexity.

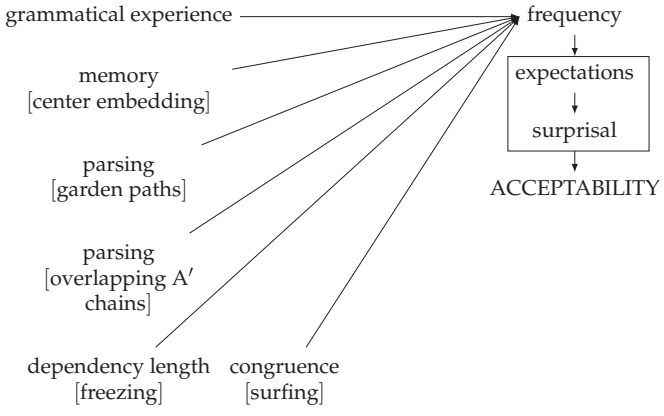


Figure 5. The logic of acceptability judgments for grammatical conditions, version 5.

4.4. Initial Non-Subjects in Zero-Relatives

Another way in which an A′ chain might incur processing complexity is if speakers are unable to infer an appropriate structure on the basis of the cues provided by the overt string in which the chain is realized—i.e., if the string associated with the otherwise well-formed A′ dependency lacks the kinds of overt signals that the processor relies on in order to parse correctly. A plausible instance of this is Jackendoff and Culicover's (1972) example in (5). This is also what happens in some instances of zero-relatives, as explored in Culicover (2013a). Consider the three relative clauses in (17).

(17) *War and Peace* is
- a. a book which you should read.
- b. a book that you should read.
- c. a book ∅ you should read.

These examples show that a relative may be introduced by a *wh*-form, *that* or zero (∅). What we see in (18) is that an initial non-subject can occur in the first two, but not the zero-relative.

(18) *War and Peace* is
- a. a book which if you have time you should read.
- b. a book that if you have time you should read.
- c. * a book ∅ if you have time you should read.

Culicover (2013a) shows that the unacceptability seen in (18c) can arise in a number of other ways, as well. In (19a) there is an initial topicalized argument,[16] in (19b) there is a initial negative constituent that triggers subject-aux-inversion, and in (19c) there is an initial predicate and stylistic inversion.

(19)
- a. * He is a man liberty$_j$, we could never grant t_j to t_i. (Cf. ?He is a man that$_i$ liberty$_j$, we could never grant t_j to t_i. (Baltin 1981)
- b. * He is a man under no circumstances would I give any money to t_i. (Cf. He is a man that$_i$ under no circumstances would I give any money to t_i)
- c. * Detroit is a town in almost every garage can be found a car manufactured by GM. (Cf. Detroit is a town that in almost every garage can be found a car manufactured by GM.)

These, along with (18), illustrate four different constructions, with the initial constituent attached to a different position in the structure. The initial subordinate clause is very high up in the structure, and can be followed by a topicalized argument, as in (20).

(20) If you have time to read a book, *War and Peace* you should definitely read.

The initial negative constituent may follow a topicalized argument, and would therefore appear to be attached lower.

(21) To Sandy, not a single dollar would I give!

The initial predicate is arguably in Spec,IP, the conventional subject position (Culicover and Levine 2001).

Thus, there does not appear to be a single syntactic configuration that could be identified in a single syntactic constraint that accounts for the unacceptability of all of these cases. Given the diversity of syntactic configurations observed here, there would have to be a separate constraint for each case, which is clearly not an optimal account. There is a common factor, however: there is a non-subject or non-NP subject in the initial position in the zero-relative clause. As a consequence, in a zero-relative there is no reliable marker of the initial portion of the relative clause. As Culicover (2013a) argues, while zero-relatives with initial NP subjects are quite standard, non-NPs in initial position in relatives are rare. Thus, when the complementizer *that* is absent and there is a non-subject or non-NP in initial position, the processor has no way of reliably identifying and projecting the relative clause structure. We suggest that the unacceptability of topicalization in zero-relative clauses reflects processing complexity, not a set of grammatical constraints.

The factor at play in this case has to do with the prediction of syntactic structure in the course of processing. As suggested in the parsing literature (Hale 2001, 2003; Levy 2005, 2008, 2013; Levy and Jaeger 2007), the sentence processor makes predictions about the future trajectory of the parse based on frequency. In this case, surprisal reflects how expected a particular syntactic category is on the basis of the structure that has already been built. This notion of expectation covers cases such as certain garden paths, where on the basis of the currently parsed string—the PREFIX—the immediately processed constituent is strongly unexpected, leading to high surprisal. An example is (22), where the prefix *without her* creates the expectation that *her contributions* is the NP complement of the preposition.[17]

(22) Without her contributions failed to come in.
(Pritchett 1988, p. 543)

There is strong evidence that the human sentence processor is continuously engaged in predicting words and structures (for a recent review, see Kuperberg and Jaeger 2016). A plausible model of such prediction is one in which at any point in the process, every possible well-formed continuation of the prefix is assigned a probability that reflects its likelihood (van Schijndel et al. 2013). In cases where the actual continuation deviates radically from what is most probable, a garden path occurs. However, when the flux of expectation is not dramatic, there is still variation in processing activity due to surprisal (Shain et al. 2020). It appears, therefore, that unacceptability judgments can be associated with levels of surprisal that exceed some threshold. (See Fodor's (1983, p. 190) discussion of "markedness" in GPSG parsing and Ross's (1987, p. 310) discussion of the accumulation of "losses in viability" for early proposals along these lines.)

In Figure 6 we add prediction of structure to the list of factors.

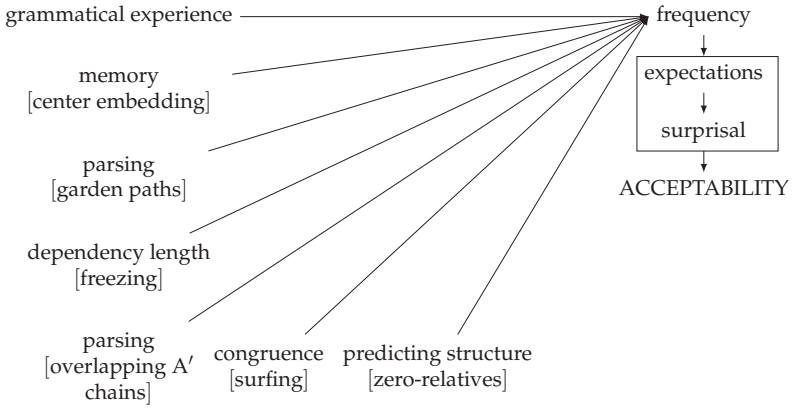


Figure 6. The logic of acceptability judgments for grammatical conditions, version 6.

5. Discourse and Information Structure

In order to communicate efficiently, speakers must provide hearers with just enough information to meet the particular goals of their conversational interaction (Grice 1975), which implies, among other things, identifying the referents the discourse is about (Ariel 1990, 2001, 2004; Roberts 2012). Excessive or irrelevant information leads to redundancy and puts the hearer through unnecessary effort, which increases processing complexity. Too little information leads to ambiguity, which also increases complexity. The management of several discourse referents at once can also lead to processing dififculties (Arnold and Griffin 2007; Gibson 2000; Kluender 2004; Warren and Gibson 2002). We argue below that these factors are plausible sources for the surprisal effect in several unacceptable A′ extractions.

5.1. Weak Crossover

We start by looking at phenomena that have to do with the computation of reference in discourse. Notably, the relevance of discourse reference to phenomena covered by syntactic constraints was already argued for in detail by Kluender (1998).

The first phenomenon, weak crossover (WCO), is exemplified by the unacceptability of examples such as (23b), first observed by Postal (1971).

(23) a. Who$_i$ t_i loves his$_i$ dog?
b. * Who$_i$ does his$_i$ dog love t_i?

Culicover (2013d), using data from Levine and Hukari (2006), argued that WCO violations such as (23b) do not reflect a principle of grammar. While the first such principle to be proposed was the Bijection Principle of Koopman and Sportiche (1983), the point is a general one: unacceptability of WCO shows the effect of referentiality and resolution of thematic assignment of chains in processing the linear string, not a syntactic constraint.[18]

Several extragrammatical factors appear to be at play. One factor is the discourse accessibility of the *wh*-phrase. The more specific the reference of the *wh*-phrase is, the more natural it is to refer to it with a pronoun, as seen in (24).

(24) a. * Who$_i$ did his$_i$ dean publicly denounce t_i?
b. ?? Which professor$_i$ did his$_i$ dean publicly denounce t_i?
c. ? [Which distinguished molecular biologist that I used to work with]$_i$ did his$_i$ dean publicly denounce t_i?

Moreover, as has been often noted, the WCO configuration with a relative clause is reliably more acceptable than precisely the same configuration with a question. Compare (25) with (24b).

(25) I plan to interview the professor who$_i$ his$_i$ dean publicly denounced t_i.

And an appositive relative is if anything even more acceptable (Lasnik and Stowell 1991); cf. (26).

(26) I plan to interview Professor Smith$_i$, who$_i$ his$_i$ dean publicly denounced t_i.

This difference can be understood in terms of specificity as well, insofar as a the head of the relative clause provides more specific information about the identity of the referent associated with the pronoun (Pesetsky 1987, 2000; Wasow 1979). The question, of course, is why this should be the case.

A second factor is whether the *wh*-phrase has a θ-role at the point in the processing of the sentence at which the pronoun is encountered. In (27a,c) the *wh*-phrase lacks a θ-role at the pronoun in the first conjunct, which contains the bound pronoun. However, in (27b,d) the *wh*-phrase gets a θ-role in the first conjunct and the pronoun is in the second conjunct.

(27) a. ? Who$_i$ does his$_i$ mother love t_i and Sandy dislike t_i?
b. Who$_i$ does Sandy dislike t_i and his$_i$ mother love t_i?
c. ? a person who$_i$ his$_i$ mother loves t_i but Sandy dislikes t_i
d. a person who$_i$ Sandy dislikes t_i but his$_i$ mother loves t_i

While the examples with the WCO violation in the first clause are somewhat marginal, those with WCO in the second clause are unobjectionable. Again, the question is why.

These factors are reducible to the degree of ACCESSIBILITY of the discourse representation corresponding to the *wh*-phrase at the point where the pronoun is encountered.[19] Accessibility is understood as a property of non-linguistic mental representations that determines their ease of retrieval in real-time processing (Arnold 2010). In the case of discourse referents, accessibility is plausibly a consequence of *predictability*: i.e., a referent is more accessible to the extent that it is more likely to be mentioned in the context at hand (Arnold 2010; Arnold and Tanenhaus 2011; Givón 1983).

We noted above that economy in referential processing seems to favor an inverse correlation between the accessibility of a discourse referent and the amount of information conveyed by the expression used to refer to it. As a result, less informative NPs (e.g., pronouns) are optimal candidates for retrieving highly accessible referents and more informative NPs (e.g., names, definite descriptions) are optimal candidates for retrieving less accessible referents (Almor 1999, 2000; Almor and Nair 2007; Ariel 1988, 1990, 1991, 1994, 2001, 2004). Different types of NP function, thus, as specialized markers for different degrees of accessibility. Whenever speakers fail to match their choice of NPs to the degree of accessibility of the referent they intend to pick out, processing complexity ensues.

Personal pronouns, like the ones we see in the WCO examples, function as HIGH ACCESSIBILITY MARKERS – i.e., they must be paired with discourse referents that are highly accessible in the contexts where they appear. This occurs because pronouns are informationally impoverished; the only kind of information pronouns carry is their specification for features such as number, person, and gender (Almor 2000; Almor and Nair 2007; Ariel 2001; Bouchard 1984; Gundel et al. 1993; Levinson 1987, 1991).

As an illustration consider the example in (28):

(28) Charlie and Frank finished watching a movie. Charlie was the one who picked it out. He didn't like it.

The personal pronoun *he* can successfully refer to Charlie in (28), because Charlie is a unique and highly accessible referent at the point where the pronominal is encountered. The discourse referent anchored to Frank is much less accessible, and, therefore, it would be odd for a speaker to use an uninformative form like *he* to refer to Frank in that context.[20]

We suggest that this same factor contributes to the unacceptability of typical WCO structures like (23b): the referent of the *wh*-phrase is not accessible enough to be retrievable

by a high accessibility marker such as a pronoun at the point where the pronoun is encountered. The mismatch between the low degree of accessibility of the discourse referent and the high accessibility marking status of pronouns contributes to processing complexity (Almor 1999, 2000). This leads to lower frequency of the WCO configurations in speakers' prior experience, which, in turn, yields higher levels of surprisal.

Gernsbacher (1989) showed in a series of experiments how various linguistic factors may enhance the relative accessibility of discourse referents. One of her major findings is that more explicit expressions (i.e., low accessibility markers, in the sense of Ariel 1990) increase the accessibility of their mental representations more than less explicit expressions. In fact, as Ariel (2001, p. 68) points out, there is an inverse relationship between an NP's degree of accessibility marking and its potential to boost the future accessibility of its discourse referent: "the lower the accessibility marker used, the more enhanced the discourse entity coded by it will become".

What all of the amelioration effects in (24)-(27) share is that they increase the accessibility of the discourse representation corresponding to the *wh*-phrase in precisely this way. When the discourse representation of the *wh*-phrase becomes more accessible, subsequent retrieval by a high accessibility marker such as a pronoun becomes more acceptable. For example, in (27), we may think of the θ-role as contributing to more information about the referent of the *wh*-phrase, which, in turn, enhances the accessibility of the mental representation it corresponds to. Increasing specificity has a similar effect in (24b,c), (25), and (26). By providing a more adequate match between the accessibility status of the antecedent and the pronoun, these ameliorated WCO violations are less complex than the unacceptable cases. They are, therefore, expected to be more frequent and to be associated with lower degrees of surprisal, enhancing acceptability.

In Figure 7 we add discourse accessibility to the list of factors.

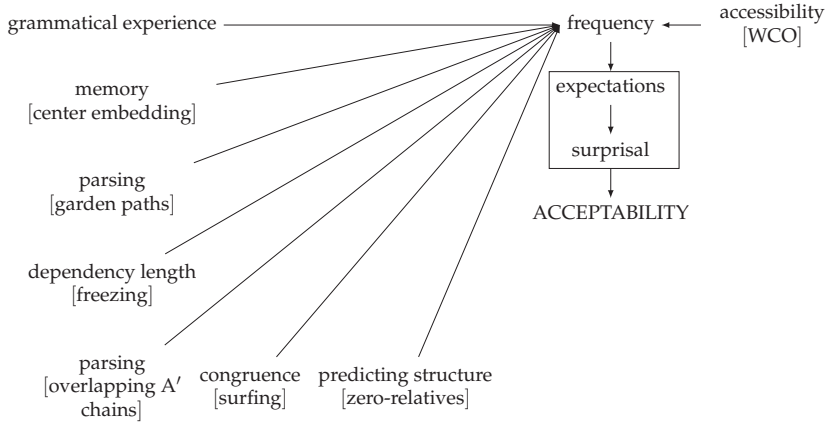


Figure 7. The logic of acceptability judgments for grammatical conditions, version 7.

5.2. The Uninvited Guest

We consider next the role that referential processing plays in the unacceptability of extraction from subjects, which conventionally falls under the Sentential Subject Constraint of Ross (1967), the Subject Condition of Chomsky (1973), and related formulations. The examples in (29) illustrate:

(29) a. * a person who (not) shaking hands with *t* would really bother Sandy (gerund)
b. * a person who us shaking hands with *t* would bother Sandy (gerund with pronominal subject)
c. * a person who Terry shaking hands with *t* would bother Sandy (gerund with referential subject)

d. * a person who Terry's shaking hands with *t* would bother Sandy (gerund with possessive)
e. * a person who that Terry shakes hands with *t* would bother everyone (*that* clause)
f. * a person who to shake hands with *t* would bother Sandy (infinitive)
g. * a person who for { us / Terry } to shake hands with *t* would bother Sandy (*for-to* infinitive)
h. * a person who offensive jokes about *t* would bother Sandy? (NP)
i. * a person who the fact that Sandy shakes hand with *t* would bother Terry (sentential complement of N like *belief*, *claim*)

In spite of the unacceptability of examples such as these, there is a substantial literature that demonstrates that the extraction from subjects is grammatical and varies in acceptability according to a number of factors, including lexical choice and repeated exposure (Abeillé et al. 2020; Chaves 2013; Chaves and Dery 2019; Chaves and Putnam 2020; Kluender 2004; Polinsky et al. 2013). Culicover and Winkler (2022) argue that in some cases, the unacceptability of extraction from subjects reflects the complexity of such extraction combined with a novel referential expression in the predicate, which they call the UNINVITED GUEST. In (29a–c), for example, the Uninvited Guest is *Sandy*.

In terms of the general model in Figures 4–7, when the complexity of a subject extraction is coupled with the complexity afforded by having to process an additional referential argument, we get a more complex and, therefore, less frequent structure, which carries a high degree of surprisal. On the account proposed by Culicover and Winkler (2022), the amelioration effect we see in connection to parasitic gaps is a consequence of reducing complexity in referential processing by omitting an extra referential argument (the Uninvited Guest). This effect can be seen in (30a–c), compared with (29a–c). The notation *pg* indicates a parasitic gap.

(30) a. a person who (not) shaking hands with *pg* would bother *t*
b. ? a person who us shaking hands with *pg* would bother *t*
c. ? a person who Terry shaking hands with *pg* would bother *t*[21]
d. * a person who Terry's shaking hands with *pg* would bother *t*
e. * a person who that Terry shakes hands with *pg* would bother *t*
f. * a person who to shake hands with *pg* would bother *t*
g. * a person who for { us / Terry } to shake hands with *pg* would bother *t*
h. * a person who the fact that Sandy shakes hands with *pg* would never bother *t*

The fact that the parasitic gap configuration is not sufficient to render all of these extractions from subjects acceptable suggests that the unacceptability here is not a matter of grammaticality per se. It is simply not the case that the presence of an extra gap elsewhere in the sentence provides a syntactic means to make subject extractions automatically grammatical, as proposed in the syntactic theories of parasitic gaps (Chomsky 1986; Frampton 1990).

This observation is further supported by the fact that there are many acceptable extractions from subject in corpora in sentences that do not contain an extra gap that could syntactically license the gap within the subject. A few examples are given in (31).

(31) a. . . . with them—the people who love you and who you love, who you laugh with and **who spending time with** is enriching rather than exhausting.
b. More than anything though, The Joker is a fascinating character **who spending time with** is a treat.

c. There are some things **which fighting against** is not worth the effort. Concentrating on things which can create significant positive change is much more fruitful.

d. That might be a good idea, the only way I could get her contact information would be through my SM though, **which asking for** would become a fiasco.

Chomsky (2008) attributes the difference in acceptability of extraction from subject to the underlying position of the subject. On Chomsky's account, if an NP is the underlying complement of a verb, extraction from subject is possible, but if it is an underlying subject, it is not. Passives would all be of the first type, as would unaccusatives, while unergatives would be of the second type. In this way, Chomsky preserves the view that subjects are islands in the grammatical sense.

However, Culicover and Winkler (2022) provide corpus evidence that extraction from subject may be acceptable even if the predicate is transitive, if the NP in the predicate does not denote a novel discourse referent, that is, if it is not an Uninvited Guest. This NP is an 'Invited Guest'—the discourse referent it invokes is highly accessible in the discourse context, implying that it carries less cost for referential processing. In every instance, the Invited Guest that has the discourse status 'given' or 'c-construable' (Rochemont and Culicover 1990), is by virtue of being part of the common ground.

A sample of Invited Guest examples is given in (32)–(36). When the object NP refers to an individual, that individual is always immediately available in the discourse, i.e., the speaker (32), the addressee or generic *you* (33), or a third party who is being discussed (34). Where the object NP does not refer to a person, it typically refers to a property of the general common ground such as *the day*, *my life* (35). The only apparent exceptions are *your playing*, *your patriotism* and *the postulated meaning* in (36), which bear on the topics of the discourse, and therefore have the discourse status 'given'.

(32) First person

a. I've found people **who spending time with** isn't an exhausting experience and actually gives me a boost.

b. However, there have been girls **who spending time with** and going places [sic] because we love them have made us happy.

(33) Second person

a. In your head you're able to let the mind wander to all sorts of corners, day dreaming about the happy things you hope might happen one day, the good times you've enjoyed, and the people **who spending time with** makes you feel good.

b. there are some people **who talking to** gives you a sort of high

c. … Deathstroke, and some other important characters, such as Alfred (**who talking to** gives you more …), James Gordon, and Barbara Gordon.

d. The purpose of a relationship (in my mind) is to find someone **who spending time with** makes you happier than you would be on your own, this guy's behaviour does not represent that in my opinion and it certainly doesnt sound like minor character traits that you may be able to change with time because it doesn't sound like he's at all willing to change.

(34) Third person

a. But even if that were so, it would seem that he had at least one person in his life **who spending time with** and whose love made him feel pure bliss.

b. … But there was one part of Tim **which to describe as typical** rather undersells him, although it is an aspect of his being to which we would all aspire, because Tim's integrity—his sense of honour, his honesty, his deep sense of decency—was special and it was rare.

c. Until Marinette, the shy classmate who tended to word-vomit in his vicinity and otherwise cease being able to function like a normal human for reasons he had yet to understand (and **which asking about** would get him sly looks from Alya and concerned looks from Nino), was there.

(35) Common attribute

a. Do you have vendors you work with that you truly enjoy? People who work hard for you, do a great job and **who spending time with** makes the day go by happily and productively?

b. Today, there was this person **who talking to** would make my life exponentially more complicated and fucked up.

(36) Sentence topic

a. Definitely the most important advice is to join an orchestra. You will not only meet likeminded individuals **who spending time with** will improve your playing, but friends and connections for life.

b. I desire that you accept of no offers of transportation from officials who deprived you of the very food, in some cases, which was necessary to supply your pressing wants, and who couple their offers of a free passage with conditions **which to accept** would cast a stain upon your patriotism as Irishmen and as free citizens, who are bound to sympathize with every struggling nationality.

c. For purposes of Proof the important distinction lies solely between assertions capable of denial with a meaning, and those **which to deny** would contradict the postulated meaning.

The data presented by Culicover and Winkler (2022) thus supports the position that there is no grammatical constraint that blocks extraction from subjects. Rather, the extraction varies in unacceptability due to a number of factors ultimately related to referential processing. When the extraction is marginally acceptable and the Uninvited Guest is absent, acceptability associated with parasitic gaps results. However, when the Uninvited Guest is present, it adds complexity to existing complexity, resulting in a judgment of unacceptability.

The Uninvited Guest analysis adds support to the claim that there does not appear to be strong evidence that non-local unacceptability in these cases is due to a grammatical constraint, although the question of why extraction from subject is complex remains open. One possible answer is that neither the *wh*-phrase nor the subject have a θ-role at the point at which the trace of the *wh*-phrase is encountered. We already saw in the case of WCO that interpreting an unresolved *wh*-chain appears to be relatively costly. Furthermore, Frazier and Clifton (1989); Kluender (2004); Kluender and Kutas (1993a) provided experimental evidence that initiating processing of an embedded sentence has a processing cost. Gibson (1998) showed that processing of referential expressions, including reference to specific times, has a cost when a *wh*-chain is not resolved. Thus it is not surprising that the most acceptable extraction from subject is from a gerund such as *shaking hands with NP*, less acceptable extraction is from a gerund with a subject such as *Terry shaking hands with NP*, and still less acceptable extraction is from a tensed S such as *that Terry shakes hands with NP*.

Figure 8 adds the processing of discourse referents to the list of factors.

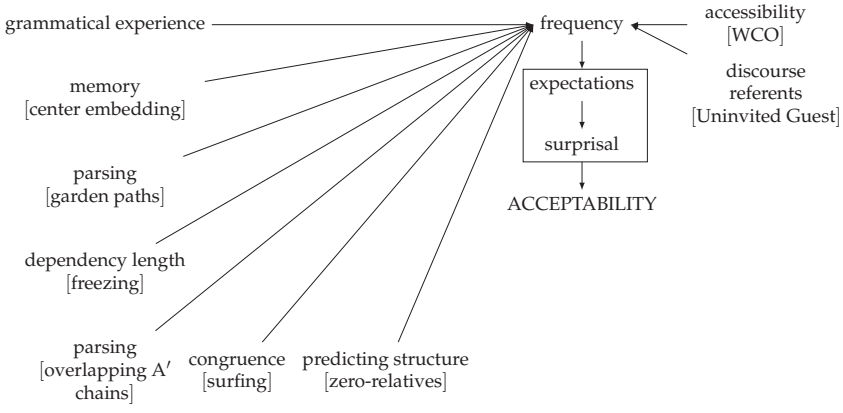


Figure 8. The logic of acceptability judgments for grammatical conditions, version 8.

5.3. Information Structure

One of the reasons to extend the RUH beyond sentence processing in a narrow sense is that there are cases in which it appears that an information structure mismatch contributes to judgments of unacceptability. The mis-management of information flow is, of course, also connected to processing complexity in a more holistic sense, having to do with the discourse as a whole.

Discrepancies between the at-issue content of utterances and the Question Under Discussion (QUD), as Roberts (2012) describes them, can cause processing difficulties (De Kuthy and Konietzko 2019; Konietzko et al. 2019). To take a simple example, the sentence (37b), while well-formed, is an inappropriate answer to the question preceding it, which functions as the QUD in that particular context. (Capitalization marks prosodic accent (focus).)

(37) Who ate the pizza?
  a. SANDY ate the pizza.
  b. # Sandy ate the PASTA.

It is likely that such mismatches fall under the general category of surprisal, but whether they recruit the same resources as garden paths and other cases that involve structure as well as interpretation is an open question.

There is evidence that information structure mismatches of this sort also play a role in acceptability judgments in extraction constructions. We cite two studies that demonstrate this. First, Culicover and Winkler (2018), following Winkler et al. (2016), observe that the acceptability of extraction from the German *was-für* construction is higher if extraction is from a focus. Compare the examples in (38)/(39), due to Müller (2010, p. 61(36)).

(38) *Was$_i$ haben [$_{DP}$ t_i für Bücher] [$_{DP}$ den FRITZ] beeindruckt?
what have [$_{DP}$ t for books.NOM] [$_{DP}$ the Fritz.ACC] impressed
'What kind of books impressed Fritz?'

(39) Was$_i$ haben [$_{DP}$ den Fritz]$_j$ [$_{DP}$ t_i für BÜCHer] t_j beeindruckt?
what have [$_{DP}$ the Fritz.ACC] [$_{DP}$ t for books.NOM] t impressed
'What kind of books impressed Fritz'

On Müller's account, *was für Bücher* in (38) is frozen, because it is last-merged in the specifier-position of vP, and hence blocks extraction. However, it is not frozen in (39), because the movement of *den Fritz* over it by scrambling removes the offending configuration that froze it—this is what Müller calls 'melting'.

However, Winkler et al. (2016) note that in German, the immediate preverbal position is a focus position (Haider and Rosengren 2003; Höhle 1982; Reis 1993; Selkirk 2011; Truckenbrodt 1995, among others). Extraction from focus in the German Mittlefeld has been independently shown to be more acceptable than extraction from non-focus (Bayer 2004). Thus, (38) is unacceptable because *Bücher* is not a focus, while (39) is more acceptable. They show that judgments of extraction from immediate preverbal and scrambled position can be manipulated by changing the context to change the focus, which rules out an explanation in structural terms.

Second, Konietzko (forthcoming) explores in detail PP extraction from subjects in German. He shows that such extraction is also sensitive to information structure and context—extraction from a focus is more acceptable than extraction from a non-focus. Konietzko shows as well that PP extraction from NP in German is sensitive to the argument type of the NP. Extraction from unaccusative subjects is best, followed by unergative subjects, transitive objects, and transitive subjects. A summary of Konietzko's results for extraction of *von wem* 'by whom' appears in Figure 9.

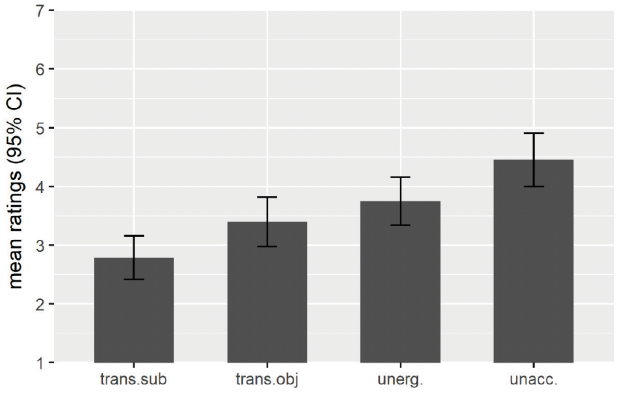


Figure 9. Acceptability of extraction of *von wem* 'by whom' from NP in German (Konietzko forthcoming).

Extraction of *über wen* 'about whom' from an NP shows sensitivity to the argument type as well (Figure 10) . Most acceptable is extraction from the subject of a passive, followed by subject of unaccusative, transitive, and psych-verb. The differences between these types of subjects have been dealt with in mainstream generative grammar in derivational terms. Konietzko concludes that there is a basis for attributing the unacceptability of at least some cases of extraction from subject to structural configuration.

Note that *wh*-constituents are canonically associated with the status of discourse *foci* (Culicover and Rochemont 1983). What happens in (38)–(39) as well as in the cases of PP extraction from subjects examined by Konietzko (forthcoming), is that full acceptability only occurs if the focus implied by the *wh*-construction is coherent with the focus associated with the structural position from which extraction takes place (the immediate pre-verbal position in (39)).

What we see in (38) is a non-optimal alignment between the information structure status of the *wh*-phrase and *den Fritz*, both of which are assumed to be foci by default. The suggestion of multiple conflicting foci arguably makes the example harder to process than (39). As a result, structures like (38) are expected to be less frequent, to give rise to higher surprisal and, correspondingly, lower acceptability.

Based on the observations in this section and Section 4, we complete our picture of the sources of unacceptability in Figure 11.

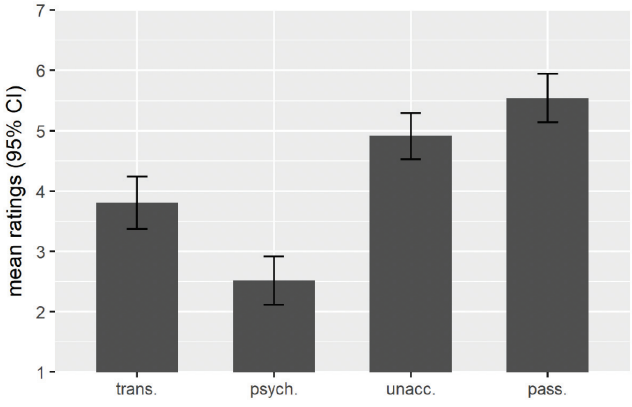


Figure 10. Acceptability of extraction of *über wen* 'about whom' from NP in German (Konietzko forthcoming).

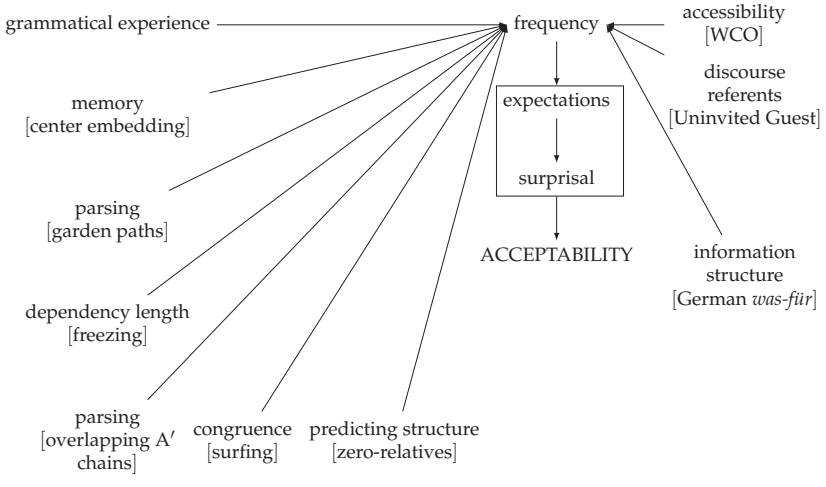


Figure 11. The logic of acceptability judgments for grammatical conditions, final version.

6. Processing Factors and Problematic Cases

As mentioned above, there are some classical island constraints that do not seem to be so readily amenable to a non-syntactic treatment. In this section, we examine specifically the Coordinate Structure Constraint and the Left Branch Condition. The phenomena covered by these constraints are prima facie counterexamples to the strongest interpretation of our hypothesis. We argue that, while there are still many open questions, there is suggestive evidence that these principles are still compatible with the ERUH.

We start by noting that it is possible that the grammar itself is a source of low frequency in a way that does not imply the existence of non-local constraints. A plausible case for this can be made for the Coordinate Structure Constraint, stated in (40) in a form that incorporates the familiar across-the-board (ATB) exceptions:

(40) *Coordinate Structure Constraint* (Ross 1967, p. 89)
In a coordinate structure, (a) no conjunct may be moved, (b) nor may any element contained in a conjunct be moved out of that conjunct unless the same element is moved out of both conjuncts.

Following previous work (Grosu 1973; Oda 2017; Pollard and Sag 1994), we distinguish between the CONJUNCT CONSTRAINT (clause (a) in (40)) and the ELEMENT CONSTRAINT (clause (b) in (40)). The former is illustrated in (41a) and the latter in (41b):

(41) a. * Who$_i$ did you see [Joanne and t_i] yesterday.
b. * Which book$_i$ did you say that [Amy wrote t_i and Harry bought the magazine]?

Though there are numerous counter-examples to (40b) which suggest that it might be reduced to a discourse-level principle (Goldsmith 1985; Kehler 1996; Kubota and Lee 2015; Lakoff 1986), (40a) seems to be a solid generalization about how coordination works in various languages (Chaves and Putnam 2020).[22]

There are, however, several alternative explanations for the robust effect illustrated in (41a) that do not involve a non-local grammatical constraint on extraction. As many authors point out, this effect follows automatically from two independently motivated proposals: the traditional analysis of coordinating conjunctions as non-heads (Bloomfield 1933; Borsley 2005; Chaves 2007; Gazdar 1980; Gazdar et al. 1985; Pesetsky 1982; Ross 1967) and the traceless account of filler-gap dependencies that is the hallmark of HPSG since the mid-1990s (Bouma et al. 2001; Chaves 2020; Chaves and Putnam 2020; Ginzburg and Sag 2000; Pickering and Barry 1991; Pollard and Sag 1994; Sag 1997; Sag and Fodor 1994).

The former idea is motivated by the basic observation that the distribution of a coordinate phrase is mainly determined by that of its conjuncts (a conjunction of NPs functions like an NP, a conjunction of VPs functions like a VP, etc.). The latter idea, in turn, is based on the hypothesis that unbounded dependency gaps are introduced by heads, rather than by phonologically null constituents (i.e., traces). This proposal requires a lexical rule which allows a head to omit one of its arguments from surface realization while at the same time introducing a corresponding gap in its argument structure. The general point is the following: If A′ gaps are not syntactic constituents, but are licensed as syntactically unrealized arguments of a head *via* a lexical rule, then coordinating conjunctions, *qua* non-heads, will not be able to co-occur with gaps.

An alternative account of the Conjunct Constraint that does not presuppose a traceless theory of extraction is suggested by Levine (2017, pp. 317–18) and Kubota and Levine (2020, pp. 302–3). They argue that the effects of (40a) can be derived from a prosodic restriction on coordinate structures requiring that each coordinated phrase contain at least one stressed syllable (see also Zwicky 1986). This is motivated by the observation that phonologically reduced cliticized pronouns cannot occur in coordinations like (42):

(42) I don't know what happened to Taylor, but it's been years since I heard from Sandy $\left\{ \begin{array}{c} \text{or him} \\ \text{*or 'm} \end{array} \right\}$.

Since extraction gaps are never phonologically realized, they cannot bear stress on their own. Therefore, in the context of NP coordinations, they cannot avoid violating this prosodic constraint.

Regardless of which theory is ultimately correct, both traceless and prosodic accounts derive the empirically robust part of the Coordinate Structure Constraint without appealing to a non-local grammatical constraint. These accounts explain the effects of (40) by means of what amount to LWFCs, thereby preserving the ERUH. The traceless theory appeals to the nature of the rule that establishes extraction gaps and the prosodic account appeals to a constraint on the prosody of the local sisters of coordinators.

Another of the classical island constraints that has resisted analysis as a consequence of non-syntactic factors is Ross's (1967) Left Branch Condition, stated in (43) and illustrated in (44):

(43) *Left Branch Condition (LBC)* (Ross 1967, p. 207)
No NP which is the leftmost constituent of a larger NP can be reordered out of this NP by a transformational rule.

(44) a. * Whose$_i$ did you read [$_{NP}$ t_i book]?
b. * His$_i$, I don't think you liked [$_{NP}$ t_i food].
c. * How much$_i$ did she earn [$_{NP}$ t_i money].

In Ross's formulation, the LBC blocks the extraction of the left branch of an NP, and requires that the phrase be pied-piped. Ross also noted that the LBC appears to be more general, and extends to examples such as (45). On the basis of such cases, Gazdar (1981) formulated a GENERALIZED LEFT BRANCH CONDITION, whose purpose is to block extraction of *any* element to the left of a lexical head (see also Emonds 1985).

(45) a. * How$_i$ is Sandy [$_{AP}$ t_i tall]? (Cf. How tall is Sandy?)
b. * [How big]$_i$ did you buy [$_{AP}$ t_i a house]? (Cf. How big a house did you buy?)

Chaves and Putnam (2020, pp. 196–200) point out that their traceless account of movement also derives these effects. If gaps do not originate as traces, but on the argument structure of heads, elements that cannot be construed as arguments of a head (determiners and other pre-nominal specifiers), will not be able to appear as gaps—i.e., they are predicted to be unextractable.

This strategy of using the rule that introduces gaps to derive the LBC faces challenges. Chief among these is the fact that, as Ross (1967 pp. 236–38) himself recognized, there are counterexamples to even the more restrictive statement of the LBC in (43) in languages like Russian and Latin.

(46) Čuju$_i$ ty čitaješ [$_{NP}$ t_i knigu]?
whose you read book
'Whose book are you reading?'

(47) Cuius$_i$ legis [$_{NP}$ t_i librum]?
whose read.2SG book
'Whose book are you reading?'

The fact that the LBC can be systematically violated in some languages suggests that it should be handled with a different strategy from the Conjunct Constraint, which is basically exceptionless. In particular, we certainly do not want to derive it from the very mechanism that builds A′ chains like Chaves and Putnam (2020) do, as this would either make wrong predictions about (46) and (47) or force us to adopt otherwise unmotivated structures for these languages.[23]

Thus, in spite of the robustness of the LBC, there are reasons to think, with Ross, that it is not a universal constraint on extraction. There is additional evidence to support this hypothesis. First, as has been recognized for some time, extraction of a subject (widely thought of as a left branch position) is acceptable in English (48).

(48) Who$_i$ do you believe [$_S$ t_i will win]?

As Grosu (1974, p. 309) observes, extraction of a possessive NP is impossible even when it is not on a left branch, as in (49) (compare with (44a)).

(49) a. * Whose$_i$ did you read [$_{NP}$ some books of t_i]? (Cf. You read some books of Susan's.)
b. * Your wife's$_i$, I met [$_{NP}$ an uncle of t_i]. (Cf. I met an uncle of your wife's.)

These last examples suggest that the problem is not with left branch extraction per se. It is reasonable to conclude, then, that there is no grammatical constraint along the lines of Ross's LBC or its generalized variant.

The explanation for the ungrammatical examples in (44)–(49) remains unclear, of course. That said, the ungrammaticality of (49a) and (49b) suggests that the problem is that the A′ constituent is by default processed as a phrasal argument with an elided

nominal head, e.g., [$_{NP}$ *whose* [$_{N}$ ∅]]. Such an analysis renders cases such as (44a) and (49a) unparseable, since there is no suitable gap for the A′ chain and no suitable parse of the NPs [*t book*] and [*some books of t*]. Something similar plausibly applies to the other cases: e.g., there is a tendency to parse the displaced constituent at the left edge of the NP in (44b) as [$_{NP}$ *his* [$_{N}$ ∅]] (as in *I liked most of the food they brought to the party, but his$_i$ I did not like t$_i$*), and, in (44c), as [$_{NP}$ *how much* [$_{N}$ ∅]] (as in *How much$_i$ did she earn t$_i$?*).

The general principle at work here seems to be a preference for parsing strings in A′ positions as full phrasal projections. This gives rise to a garden-path effect when the speaker encounters an NP missing a left branch. Whether this idea is on the right track, and whether it can be extended to all other cases handled by the LBC is a question that we leave open here.

7. Summary

Let us summarize. For almost every constraint on extraction that has been noted in the literature, including classical strong islands, we have suggested that it is possible to identify a plausible non-syntactic cause or causes. For the single case where a non-syntactic cause seems implausible (the Conjunct Constraint), a purely local well-formedness condition seems to be sufficient. The picture that emerges is consistent with the ERUH.

Extended Radical Unacceptability Hypothesis: All judgments of reduced acceptability in cases of otherwise well-formed (i.e., locally well-formed) extractions are due to non-syntactic factors, not syntactic constraints.

Thus, it appears that there is limited support for grammatical constraints as accounts of the unacceptability of extraction from islands. It is in fact reasonable to hypothesize that in virtually every case of unacceptability, if the local well-formedness conditions of the grammar are satisfied, the reason for the unacceptability is non-syntactic. Processing complexity appears to be the most prominent candidate, which is sensitive to syntactic configuration, discourse accessibility, pragmatic plausibility including relevance, contextual factors such as information structure, and frequency.

That said, there are several major open questions that have to be dealt with. One is to see whether our ERUH-compliant explanations for the Coordinate Structure Constraint and Left Branch Condition hold up under closer empirical scrutiny. There are also cases of apparent freezing that involve chain interactions different from the sort discussed in the English freezing cases discussed in Section 4.1. In these cases we must seek alternate sources of low frequency, which would be sufficient to account for low acceptability in the model sketched in Figure 11.

A second question concerns cross-linguistic variation: if island and similar effects are the consequence of non-syntactic factors, why do different languages reflect differences in the extent to which they show sensitivity to island constraints? Still more problematic is evidence for inter-individual variation in judgments for particular island violations (Kush et al. 2017). One would assume that non-syntactic factors would be constant across languages and individuals. In order to account for the variation, we would suggest pursuing an explanation in terms of language-specific differences in frequency in the specific constructions that show differences in acceptability judgments. Again, the key idea is that acceptability correlates with frequency.

Finally, ERUH is a very strong hypothesis—it says that there are no purely syntactic constraints that are not LWFCs. This strong localist outlook is characteristically associated with GPSG, SBCG, and variants of Categorial Grammar (Gazdar et al. 1985; Kubota and Levine 2020; Sag 2012)—theories which confine syntactic constraints to local chunks of representation of the kind that could be encoded in a single phrase-structure rule. In those cases where there is putative evidence that syntax per se is responsible for acceptability judgments in non-local dependencies (e.g., Kush et al. 2017; Phillips 2006, 2013a), we

would always want to see if it is possible to rule out all plausible aspects of processing, pragmatics and semantics as potential explanations.

As we saw throughout our discussion here, many of the phenomena that were once plausibly analyzed as requiring syntactic constraints on non-local configurations are actually better explained in terms of non-syntactic factors. We believe that this kind of approach is plausible not only for the empirical reasons we mentioned in this paper, but also for conceptual and heuristic ones. Conceptually, a theory of grammar that subscribes to the ERUH excludes a prima facie source of complexity that would impose a heavy burden on evolutionary accounts of the syntactic component of the language faculty (Berwick and Chomsky 2016; Hauser et al. 2002; Jackendoff 2002). In addition, heuristically, the questions that the ERUH raises open a fruitful avenue of cross-disciplinary dialogue between theories of linguistic representation and theories of processing and general cognitive capacities.

Author Contributions: Conceptualization, P.W.C., G.V. and S.W.; methodology, P.W.C., G.V. and S.W.; formal analysis, P.W.C., G.V. and S.W.; writing—original draft preparation, P.W.C., G.V. and S.W.; writing—review and editing, P.W.C., G.V. and S.W. All authors have read and agreed to the published version of the manuscript.

Funding: The collaborative research reported here was partially funded by the Alexander von Humboldt Stiftung and the Deutsche Forschungsgemeinschaft (DFG, German Resarch Foundation)—SFB 833—A7—Project ID 75650358.

Acknowledgments: We wish to express our profound thanks to the reviewers for the time, effort and care that they devoted to reading our paper and pointing out errors, omissions, relevant literature, and passages in need of clarification. We are very much in debt to them. Any errors that remain are solely our responsibility.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1] Subsequent syntactic theories of islands have, of course, evolved well beyond Ross's early efforts. The main thrust of the literature on islands after Ross, as far as we can see, is to derive the central results of Ross's classical syntactic account in a more principled way, often with the goal of unifying various locality conditions (see Boeckx (2012) for an overview). However, this tradition inherits from Ross and other early work like Chomsky (1973) the idea that the patterns underlying island effects are syntactic in nature (Bošković 2015; Chomsky 2008; Phillips 2013a, 2013b; Sprouse 2007, 2012a, 2012b). Our discussion here—as well as the more lengthy arguments in Kluender (1991), Goldberg (2006), Hofmeister and Sag (2010), Chaves and Putnam (2020) and Kubota and Levine (2020)—targets this basic assumption, rather than the details of specific syntactic accounts.

[2] Due to space considerations we are unable to survey every phenomenon that bears on this hypothesis. For research on a broad array of phenomena that appear to be consistent with the ERUH, see Francis (2022). Additionally, it appears plausible that the ERUH applies to other kinds of putative non-local constraints, such as Condition C and the binding of long-distance anaphors (Reinhart and Reuland 1991; Varaschin 2021; Varaschin et al. 2022). We also do not deal with weak islands such as wh-islands and negative islands, for which a range of both syntactic and semantic accounts have been proposed. For a review, see Szabolcsi and Lohndal (2017), who conclude that "it seems true beyond reasonable doubt that a substantial portion of this large [weak island—PWC, GV and SW] phenomenon is genuinely semantic in nature", and Abrusán (2014). This work suggests that weak islands are consistent with the ERUH. See also Kroch (1998) for a pragmatic account of weak islands, and Gieselman et al. (2013) for experimental evidence that the unacceptability of extraction from negative islands arises from the interaction of various processing demands.

[3] More formally, Levy (2008) defines the surprisal associated with a given linguistic expression e_n as its negative log probability conditional on all the previous expressions in the discourse and the relevant features of the extra-sentential context (written as CONTEXT):

(i) $\text{surprisal}(e_n) = -\log P(e_n \mid e_1, \ldots e_{n-1}, \text{CONTEXT})$

Our use of surprisal is different in several respects from Levy's. First, Levy (2008) defines surprisal relative to *words*. We are generalizing the notion to linguistic expressions in general, including words and phrases. Second, Levy documents the correlation between surprisal and performance measures such as reaction times, while we are focusing on the underlying processing and acceptability responses. In this respect we are following a line of research pursued by Park et al. (2021), who use surprisal to measure a deep learning language model's knowledge of syntax. They explore the extent to which a language model's surprisal score for pairs of sentences matches with standard acceptability contrasts found in textbooks. They found that "the accuracy of

BERT's acceptability judgments [i.e., the correspondence between the surprisal value assigned by the language model, BERT, and the acceptability reported in textbooks] is fairly high" (Park et al. 2021, p. 420).

4 The frequency that determines expectations is not that of sequences of strings, but, rather, of linguistic expressions, minimally construed as correspondences of phonological, syntactic, and semantic information (Goldberg 1995, 2006; Jackendoff 2002; Michaelis 2012; Sag 2012). This caveat is necessary in order to avoid the objection Chomsky (1957, pp. 15–17) raised to statistical approaches. In the context *I saw a fragile _*, the strings *bassoon* and *of* may share an equal frequency in the past linguistic experience of an English speaker (≈0). However, since the speaker independently knows that *bassoon* is a noun and *of* is a preposition and the sequence *fragile* NP is much more frequent than *fragile* P, the expectation (and, therefore, the acceptability) for the former is much higher than for the latter.

5 For instance, in order to state a syntactic restriction against multiple center-embeddings, we would need some way of counting the number of embedded clauses; in order to account for (5), we would need the syntactic constraint on A′ movement to be sensitive to the position of the gap in the linear order of the string (which contradicts the widespread assumption that transformations are structure-dependent). The very idea of syntactic constraints on unbounded dependencies also entails a non-trivial extension of the vocabulary of syntactic theory insofar as it requires ways of referring to chunks of syntactic representations of an indeterminate size, as discussed in connection to (2) above.

6 A reviewer correctly points out that in principle failure of a particular example to observe a proposed syntactic constraint could be a 'grammatical illusion' (Christensen 2016; de Dios-Flores 2019; Engelmann and Vasishth 2009; Phillips et al. 2011; Trotzke et al. 2013). Clearly, such a possibility always exists where there are differences in judgments of acceptability. However, in order to appeal to a grammatical illusion to account for the acceptability of an island violation it is important to show that doing so results in a simplification of the theory of grammar; otherwise, one can aways appeal to a grammatical illusion in order to get around any counterexample to a proposed syntactic constraint. Quite the opposite appears to be the case for islands. As Phillips (2013a, p. 54) puts it, "[n]atural language grammars would probably be simpler if there were no island constraints" . The reasons relate to the point we made above about how syntactic accounts require extending the descriptive vocabulary of grammatical theory.

7 We note that evolutionary considerations are not incompatible per se with a syntactic approach to islands. On such a view, it would be necessary to show that island effects follow from an interaction of general architectural features of the syntactic part of language that could independently be justified on evolutionary grounds. We are not aware of such a demonstration. Hauser et al. (2002) suggest an alternative view, where island constraints arise automatically from solutions to the problem of optimizing the syntactic outputs constructed by the "narrow" faculty of language to the constraints imposed by the "broad" faculty of language – i.e., the cognitive systems that the syntax interacts with. If the latter are understood to include processing systems, Hauser et al.'s (2002) hypothesis can be seen as an instance of the RUH.

8 In fact, the experiments they report demonstrate that manipulation of frequency has an effect on acceptability judgments for island extractions.

9 Ross's formulation of the constraint reflects the fact that it is not possible to extract from an extraposed relative clause, even though it is not in a configuration that would fall under the Complex NP Constraint. Thus we see right at the start the treatment of freezing as a special type of island phenomenon.

10 For other proposals that take chain interactions to result in ungrammaticality, see Chomsky's (1977) discussion of the interaction of wh-movement and tough-movement and also Fodor (1978), and Pesetsky (1982). In contrast, Collins (2005) proposes an account of the English passive that requires movement of a sub-constituent from a larger, moved constituent.

11 It should also be noted that there are phenomena where greater distance between dependent elements appears to improve acceptability (see, for example, Vasishth and Lewis 2006). Such 'anti-locality' effects suggest that there are yet other factors at play, such as predictability related to selection (Levy and Keller 2013; Rajkumar et al. 2016). Moreover, research on the processing of relative clauses in languages such as Japanese and Korean suggests that there may be a preference of extraction of subjects over objects even though the gaps corresponding to the subjects are arguably further from the head (see, for example, Nakamura and Miyamoto 2013; Ueno and Garnsey 2008). These data favor the view that dependency length should be measured in terms of complexity of branching structure, given that in head-final languages the position of subject gaps is linearly farther but hierarchically closer to the position of the filler noun.

12 The term 'surfing' is due to Sauerland (1999).

13 For completeness we note that there is a range of cases of purported freezing that do not immediately lend themselves to explanations in terms of non-syntactic factors. Among these are phenomena in German (Bayer 2018; Müller 2018), and Dutch (Corver 2018). These phenomena await a more extensive analysis than we can provide here.

14 Crossing is also seen in another type of example that fell under the freezing account of Wexler and Culicover (1980):

(i) Which table$_i$ did you put t_j on t_i [a picture of FDR$_j$] ?

15 The dependency length literature suggests that minimization of dependency length alone is not sufficient to account for structural preferences reflecting degree of congruence (Kuhlmann and Nivre 2006). Also relevant are the degree of adjacency of dependent constituents, measured by GAP DEGREE, which measures the number of discontinuities within a subtree, EDGE DEGREE, which

measures the number of intervening constituents spanned by a single edge, and the disjointness of constituents, measured by WELL-NESTEDNESS (Kuhlmann and Nivre 2006, p. 511).

16 For analyses of the relationship between topicalization and the complementizer in terms of Optimality Theory, see Pesetsky (1998) and Grimshaw (1997).

17 For a review of a range of types of garden paths, see Pritchett (1992, 1988).

18 For a comprehensive review of WCO effects and of proposals to account for WCO, see Safir (2017). Safir notes a number of cases that are more complex than (23b) that the current proposal does not address.

19 For a computational account of crossover effects in terms of linear order processing, see Shan and Barker (2006).

20 The lower accessibility of Frank would justify repeating the name *Frank* or using some other referential phrase carrying a higher degree of informativity. Repetition of *Charlie* in (28), in turn, would have been redundant and would, as a result, contribute to increase processing complexity (Gordon and Hendrick 1998).

21 We show below that the relative unacceptability of (30c–e) vs. (30a) is related to the Uninvited Guest in virtue of the presence of additional referring expressions as subjects as well as finite tense (cf. Kluender 1998).

22 Throughout most of the history of transformational grammar, the Coordinate Structure Constraint has resisted an integration into general syntactic theories of islands like the ones proposed by Chomsky (1973, 1977, 1986, 2008). However, it did play an important role in non-transformational theories like GPSG and HPSG (Gazdar 1981; Pollard and Sag 1994). More recently, minimalist accounts of both parts of (40) have been proposed which make critical use of NON-LOCAL GRAMMATICAL CONSTRAINTS, such as Chomsky's (2000) Phase Impenetrability Condition and Rizzi's (1990) Relativized Minimality (Bošković 2020; Oda 2021). Relativized minimality counts as a non-local constraint in our sense because, even in the absence of interveners, the distance between a target position and a movement trace can still be arbitrarily large. A similar observation applies to the size of the domain of a phase (i.e., the spell-out domain), from which extraction is ruled out by the Phase Impenetrability Condition (Chomsky 2000).

23 This is ultimately the strategy advocated by Chaves and Putnam (2020, pp. 102–3).

References

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204: 104293. [CrossRef] [PubMed]

Abrusán, Márta. 2014. *Weak Island Semantics*. Oxford: Oxford University Press.

Almor, Amit. 1999. Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review* 106: 748–65. [CrossRef] [PubMed]

Almor, Amit. 2000. Constraints and mechanisms in theories of anaphor processing. In *Architectures and Mechanisms for Language Processing*. Edited by Matthew W. Crocker, Martin J. Pickering and Charles Clifton. Cambridge, MA: MIT Press, pp. 341–54.

Almor, Amit, and Veena A. Nair. 2007. The form of referential expressions in discourse. *Language and Linguistics Compass* 1: 84–99. [CrossRef]

Ariel, Mira. 1988. Referring and accessibility. *Journal of linguistics* 24: 65–87. [CrossRef]

Ariel, Mira. 1990. *Accessing Noun-Phrase Antecedents*. London: Routledge.

Ariel, Mira. 1991. The function of accessibility in a theory of grammar. *Journal of Pragmatics* 16: 443–63. [CrossRef]

Ariel, Mira. 1994. Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of linguistics* 30: 3–42. [CrossRef]

Ariel, Mira. 2001. Accessibility theory: An overview. In *Text Representation: Linguistic and Psycholinguistic Aspects*. Edited by Ted. J. M. Sanders, Joost Schilperoord and Wilbert Spooren. Amsterdam: John Benjamins Publishing Company, pp. 29–87.

Ariel, Mira. 2004. Accessibility marking: Discourse functions, discourse profiles, and processing cues. *Discourse Processes* 37: 91–116. [CrossRef]

Arnold, Jennifer E., and Zenzi M. Griffin. 2007. The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language* 56: 521–36. [CrossRef]

Arnold, Jennifer E. 2010. How speakers refer: The role of accessibility. *Language and Linguistics Compass* 4: 187–203. [CrossRef]

Arnold, Jennifer E., and Michael K. Tanenhaus. 2011. Disfluency effects in comprehension: How new information can become accessible. In *The Processin and Acquisition of Reference*. Edited by Edward A. Gibson and Neal J. Pearlmutter. Cambridge, MA: MIT Press, pp. 197–217.

Arnon, Inbal, Philip Hofmeister, T. Florian Jaeger, IvanA. Sag, and Neal Snider. 2005. Rethinking superiority effects: A processing model. Paper presented at the 18th Annual CUNY Conference on Human Sentence Processing, University of Arizona, Tucson, AZ, March 2.

Baltin, Mark. 1981. Strict bounding. In *The Logical Problem of Language Acquisition*. Edited by C. Leroy Baker and John McCarthy. Cambridge, MA: MIT Press.

Bayer, Josef. 2004. Was beschränkt die Extraktion? Subjekt-Objekt vs. Topic-Fokus. In *Deutsche Syntax: Empirie und Theorie*. Edited by Franz-Josef D'Avis. Volume 46 of *Acta Universitatis Gothoburgensis*. Göteborg: Göteborger Germanistische Forschungen, pp. 233–57.

Bayer, Josef. 2018. Criteral freezing in the syntax of particles. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Jäger, Andreas Konietzko and Susanne Winkler. Berlin: De Gruyter, pp. 224–63.
Berwick, Robert C., and Noam Chomsky. 2016. *Why Only Us: Language and Evolution*. Cambridge, MA: MIT Press.
Bever, Thomas. 1970. The cognitive basis for linguistic structures. In *Cognition and the Development of Language*. Edited by John R. Hayes. New York: John Wiley & Sons, pp. 279–362.
Bloomfield, Leonard. 1933. *Language*. New York and London: Henry Holt and Co. and Allen and Unwin Ltd.
Boeckx, Cedric. 2008. Islands. *Language and Linguistics Compass* 2: 151–67. [CrossRef]
Boeckx, Cedric. 2012. *Syntactic Islands*. Cambridge: Cambridge University Press.
Borsley, Robert D. 2005. Against Conjp. *Lingua* 115: 461–82. [CrossRef]
Bošković, Željko. 2015. From the complex NP constraint to everything: On deep extractions across categories. *The Linguistic Review* 32: 603–69. [CrossRef]
Bošković, Željko. 2020. On the Coordinate Structure Constraint, across-the-board-movement, phases, and labeling. In *Recent Developments in Phase Theory*. Edited by Jeroen van Craenenbroeck, Cora Pots and Tanja Temmerman. Berlin: De Gruyter Mouton, pp. 133–82.
Bouchard, Denis. 1984. *On the Content of Empty Categories*. Dordecht: Foris.
Bouma, Gosse, Robert Malouf, and Ivan Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19: 1–65. [CrossRef]
Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82: 711–33. [CrossRef]
Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge Univeristy Press.
Bybee, Joan L., and Paul J. Hopper. 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
Chaves, Rui P. 2007. Coordinate Structures: Constraint-Based Syntax-Semantics Processing. Ph.D. thesis, Universidade de Lisboa, Lisbon, Portugal.
Chaves, Rui P. 2013. An expectation-based account of subject islands and parasitism. *Journal of Linguistics* 49: 285–327. [CrossRef]
Chaves, Rui P. 2020. Island phenomena and related matters. In *Head-Driven Phrase Structure Grammar: The Handbook*. Edited by Stefan Müller, Anne Abeillé, Robert D. Borsley and Jean-Pierre Koenig. Berlin: Language Science Press.
Chaves, Rui P., and Jeruen E. Dery. 2014. Which subject islands will the acceptability of improve with repeated exposure? In *Proceedings of the Thirty-First Meeting of the West Coast Conference on Formal Linguistics, Arizona State University, February 7–9. 2013*. Edited by Robert E. Santana-LaBarge. Somerville, MA: Cascadilla Project. pp. 96–106.
Chaves, Rui P., and Jeruen E. Dery. 2019. Frequency effects in subject islands. *Journal of linguistics* 55: 475–521. [CrossRef]
Chaves, Rui P., and Michael T. Putnam. 2020. *Unbounded Dependency Constructions: Theoretical and Experimental Perspectives*. Oxford: Oxford University Press.
Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*. Edited by Stephen Anderson and Paul Kiparsky. New York: Holt, Reinhart & Winston, pp. 232–86.
Chomsky, Noam. 1977. On *wh*-movement. In *Formal Syntax*. Edited by Peter W. Culicover, Thomas Wasow and Adrian Akmajian. New York: Academic Press, pp. 71–132.
Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.
Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Edited by Roger Martin, David Michaels and Juan Uriagereka. Cambridge, MA: MIT Press, pp. 89–156.
Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Linguistics*. Edited by Michael Kenstowicz. Cambridge, MA: MIT Press, pp. 1–52.
Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Edited by Robert Freidin, Carlos P. Otero and Maria Luisa Zubizarreta. Cambridge, MA: MIT Press, pp. 133–66.
Chomsky, Noam, Ángel J. Gallego, and Dennis Ott. 2019. Generative grammar and the faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics* 18: 229–61. [CrossRef]
Christensen, Ken Ramshøj. 2016. The dead ends of language: The (mis)interpretation of a grammatical illusion. In *Let Us Have Articles Betwixt Us: Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood*. Edited by Sten Vikner, Henrik Jørgensen and Elly van Gelderen. Aarhus: Aarhus University, pp. 129–59.
Citko, Barbara. 2014. *Phase Theory: An Introduction*. Cambridge: Cambridge University Press.
Collins, Chris. 2005. A smuggling approach to the passive in English. *Syntax* 8: 81–120. [CrossRef]
Corballis, Michael C. 2017. The evolution of language. In *APA Handbook of Comparative Psychology: Basic Concepts, Methods, Neural Substrate, and Behavior*. Edited by Josep Call, Gordon M. Burghardt, Irene M. Pepperberg, Charles T. Snowdon and Thomas R. Zentall. Washington, DC: American Psychological Association, pp. 273–97.
Corver, Norbert. 2017. Freezing effects. In *The Blackwell Companion to Syntax*, 2nd ed. Edited by Martin Everaert and Henk van Riemsdijk. Malden: Blackwell Publishing.
Corver, Norbert. 2018. The freezing points of the (Dutch) adjectival system. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Jäger, Andreas Konietzko and Susanne Winkler. Berlin: De Gruyter, pp. 143–94.

Culicover, Peter W. 2005. Squinting at Dali's Lincoln: On how to think about language. In *Proceedings of the Annual Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, vol. 41, pp. 109–28.
Culicover, Peter W. 2013a. English (zero-)relatives and the competence-performance distinction. *International Review of Pragmatics* 5: 253–70. [CrossRef]
Culicover, Peter W. 2013b. *Explaining Syntax*. Oxford: Oxford University Press.
Culicover, Peter W. 2013c. *Grammar and Complexity: Language at the Intersection of Competence and Performance*. Oxford: Oxford University Press.
Culicover, Peter W. 2013d. The role of linear order in the computation of referential dependencies. *Lingua* 136: 125–44. [CrossRef]
Culicover, Peter W. 2015. Simpler Syntax and the mind. In *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff*. Edited by Ida Toivonen, Piroska Csuri and Emile Van Der Zee. Cambridge: MIT Press, pp. 3–20.
Culicover, Peter W. 2021. *Language Change, Variation and Universals—A Constructional Approach*. Oxford: Oxford University Press.
Culicover, Peter W., and Robert D. Levine. 2001. Stylistic inversion in English: A reconsideration. *Natural Language & Linguistic Theory* 19: 283–310.
Culicover, Peter W., and Andrzej Nowak. 2002. Markedness, antisymmetry and complexity of constructions. *Linguistic Variation Yearbook* 2: 5–30. [CrossRef]
Culicover, Peter W., and Andrzej Nowak. 2003. *Dynamical Grammar: Minimalism, Acquisition and Change*. Oxford: Oxford University Press.
Culicover, Peter W., and Michael Rochemont. 1983. Stress and focus in English. *Language* 59: 123–65. [CrossRef]
Culicover, Peter W., and Susanne Winkler. 2018. Freezing, between grammar and processing. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Jäger, Andreas Konietzko and Susanne Winkler. Berlin: De Gruyter, pp. 353–86.
Culicover, Peter W., and Susanne Winkler. 2022. Parasitic gaps aren't parasitic or, the Case of the Uninvited Guest. *The Linguistic Review* 39:1–35. [CrossRef]
de Dios-Flores, Iria. 2019. Processing sentences with multiple negations: Grammatical structures that are perceived as unacceptable. *Frontiers in Psychology* 10: 2346. [CrossRef] [PubMed]
De Kuthy, Kordula, and Andreas Konietzko. 2019. Information structural constraints on PP topicalization from NPs. In *Architecture of Topic*. Edited by Valéria Molnár, Verner Egerland and Susanne. Winkler. Berlin and New York: De Gruyter, pp. 203–22.
Deane, Paul. 1991. Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics* 2: 1–63. [CrossRef]
Emonds, Joseph E. 1985. *A Unified Theory of Syntactic Categories*. Dordrecht: Foris.
Engelmann, Felix, and Shravan Vasishth. 2009. Processing grammatical and ungrammatical center embeddings in english and german: A computational model. Paper presented at Ninth International Conference on Cognitive Modeling, Manchester, UK, July 24–26. pp. 240–45.
Erteschik-Shir, Nomi. 1977. On the Nature of Island Constraints. Ph.D. thesis, MIT, Cambridge, MA, USA.
Erteschik-Shir, Nomi. 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford: Oxford University Press.
Erteschik-Shir, Nomi, and Shalom Lappin. 1979. Dominance and the functional explanation of island constraints. *Theoretical Linguistics* 6: 43–84. [CrossRef]
Fodor, Janet Dean. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9: 427–73.
Fodor, Janet Dean. 1983. Phrase structure parsing and the island constraints. *Linguistics and Philosophy* 6: 163–223. [CrossRef]
Frampton, John. 1990. Parasitic gaps and the theory of wh-chains. *Linguistic Inquiry* 21: 49–78.
Francis, Elaine J. 2022. *Gradient Acceptability and Linguistic Theory*. Oxford: Oxford University Press.
Frazier, Lyn, and Charles Clifton. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4: 93–126. [CrossRef]
Futrell, Richard, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences* 112: 10336–41. [CrossRef] [PubMed]
Gazdar, Gerald. 1980. A cross-categorial semantics for coordination. *Linguistics* 3: 407–9. [CrossRef]
Gazdar, Gerald. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12: 155–84.
Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford and Cambridge: Blackwell Publishing and Harvard University Press.
Gernsbacher, Morton. 1989. Mechanisms that improve referential access. *Cognition* 32: 99–156. [CrossRef]
Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68: 1–76. [CrossRef]
Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, Language, Brain*. Edited by Alec P. Marantz, Yasushi Miyashita and Wayne O'Neil. Cambridge, MA: MIT Press, pp. 95–126.
Gieselman, Simone, Robert Kluender, Ivano Caponigro, Yelena Fainleib, Nicholas LaCara, and Yangsook Park. 2013. Isolating processing factors in negative island contexts. *Proceedings of NELS* 41: 233–46.
Ginzburg, Jonathan, and Ivan A. Sag. 2000. *Interrogative Investigations*. Stanford: CSLI publications.
Givón, Talmy. 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins Publishing Company.
Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldsmith, John. 1985. A principled exception to the Coordinate Structure Constraint. In *Proceedings of the Chicago Linguistic Society 21*. Edited by William H. Eilfort, Paul D. Kroeber and Karen L. Peterson. pp. 133–43.
Gordon, Peter C., and Randall Hendrick. 1998. The representation and processing of coreference in discourse. *Cognitive Science* 22: 389–424. [CrossRef]
Gould, Stephen Jay, and Richard C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 205: 581–98.
Grice, H. Paul. 1975. Logic and conversation. In *Speech Acts*. Edited by Peter Cole and Jerry L. Morgan. Volume 3 of *Syntax and Semantics*. New York: Academic Press, pp. 43–58.
Grimshaw, Jane. 1997. Projection, heads, and optimality. *Linguistic Inquiry* 28: 373–422.
Grosu, Alexander. 1973. On the nonunitary nature of the Coordinate Structure Constraint. *Linguistic Inquiry* 4: 88–92.
Grosu, Alexander. 1974. On the nature of the Left Branch Condition. *Linguistic Inquiry* 5: 308–19.
Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69: 274–307. [CrossRef]
Haider, Hubert, and Inger Rosengren. 2003. Scrambling: Nontriggered chain formation in OV languages. *Journal of Germanic Linguistics* 15: 203–67. [CrossRef]
Hale, John. 2001. A probablistic Earley parser as a psycholinguistic model. Paper presented at Second Meeting of the North American Chapter of the Association for Computational Linguistics, Morristown, NJ, USA, June 1–7. Stroudsburg, PA: Association for Computational Linguistics, pp. 1–8.
Hale, John. 2003. The information conveyed by words in sentences. *Journal of Psycholinguistic Research* 32: 101–23. [CrossRef]
Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569–79. [CrossRef]
Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
Hawkins, John A. 2014. *Cross-Linguistic Variation and Efficiency*. Oxford: Oxford University Press.
Hofmeister, Philip, Peter W. Culicover, and Susanne Winkler. 2015. Effects of processing on the acceptability of frozen extraposed constituents. *Syntax* 18: 464–83. [CrossRef]
Hofmeister, Philip, T. Florian Jaeger, Inbal Arnon, Ivan Sag, and Neal Snider. 2013. The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes* 28: 48–87.
Hofmeister, Philip, T. Florian Jaeger, Ivan A. Sag, Inbal Arnon, and Neal Snider. 2007. Locality and accessibility in *wh*-questions. In *Roots: Linguistics in Search of Its Evidential Base*. Edited by Sam Featherston and Wolfgang Sternefeld. Berlin: Mouton de Gruyter, pp. 185–206.
Hofmeister, Philip, and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415. [CrossRef]
Hofmeister, Philip, Laura Staum Casasanto, and Ivan A. Sag. 2013. Islands in the grammar? Standards of evidence. In *Experimental Syntax and the Islands Debate*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 42–63.
Höhle, Tilman. 1982. Explikation für normale Betonung und normale Wortstellung. In *Satzglieder im Deutschen*. Edited by Werner Abraham. Tübingen: Gunter Narr Verlag, pp. 75–153.
Jackendoff, Ray. 1977. *X′ Syntax*. Cambridge, MA: MIT Press.
Jackendoff, Ray. 1999. Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences* 3: 272–79. [CrossRef]
Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
Jackendoff, Ray, and Peter Culicover. 1972. A reconsideration of dative movement. *Foundations of Language* 7: 397–412.
Jäger, Marion. 2018. An experimental study on freezing and topicalization in English. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Knecht, Andreas Konietzko and Susanne Winkler. Berlin and New York: Mouton de Gruyter, pp. 430–50.
Kaplan, Ronald M., and Annie Zaenen. 1995. Long-distance dependencies, constituent structure, and functional uncertainty. In *Formal Issues in Lexical-Functional Grammar*. Edited by Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III and Annie Zaenen. Stanford, CA: CSLI Publications. pp. 137–65.
Kehler, Andrew. 1996. Coherence and the coordinate structure constraint. *Annual Meeting of the Berkeley Linguistics Society* 22: 220–31. [CrossRef]
Kluender, Robert. 1991. Cognitive Constraints on Variables in Syntax. Ph.D. thesis, University of California, San Diego, La Jolla, CA, USA.
Kluender, Robert. 1992. Deriving islands constraints from principles of predication. In *Island Constraints: Theory, Acquisition and Processing*. Edited by Helen Goodluck and Michael Rochemont. Dordrecht: Kluwer, pp. 223–58.
Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In *Syntax and Semantics 29: The Limits of Syntax*. Edited by Peter W. Culicover and Louise McNally. San Diego: Academic Press, pp. 241–79.
Kluender, Robert. 2004. Are subject islands subject to a processing account? Paper presented at 23rd West Coast Conference on Formal Linguistics, University of California, Davis, CA, USA, April 23–25. Edited by Benjamin Schmeiser, Veneeta Chand, Ann Kelleher and Angelo J. Rodriguez. Somerville: Cascadilla Press. pp. 475–99.
Kluender, Robert, and Marta Kutas. 1993a. Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience* 5: 196–214. [CrossRef]

Kluender, Robert, and Marta Kutas. 1993b. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8: 573–633. [CrossRef]
Konietzko, Andreas. forthcoming. PP extraction from subject islands in German. *Glossa* 7.
Konietzko, Andreas, Janina Radó, and Susanne Winkler. 2019. Focus constraints on relative clause antecedents in sluicing. In *Information Structure and Semantic Processing*. Edited by Sam Featherston, Robin Hörnig, Sophie von Wietersheim and Susanne Winkler. Berlin and New York: Mouton de Gruyter, pp. 105–27.
Konietzko, Andreas, Susanne Winkler, and Peter W. Culicover. 2018. Heavy NP shift does not cause freezing. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 63: 454–64. [CrossRef]
Koopman, Hilda, and Dominique Sportiche. 1983. Variables and the bijection principle. *The Linguistic Review* 2: 139–60.
Kroch, Anthony. 1998. Amount quantification, referentiality, and long *wh*-movement. In *Penn Working Papers in Linguistics, Vol. 5.2*. Edited by Alexis Dimitriadis, Hikyoung Lee, Christine Moisset and Alexander Williams. Philadelphia: University of Pennsylvania.
Kubota, Yusuke, and Jungmee Lee. 2015. The coordinate structure constraint as a discourse-oriented principle: Further evidence from Japanese and Korean. *Language* 91: 642–75. [CrossRef]
Kubota, Yusuke, and Robert D. Levine. 2020. *Type-Logical Syntax*. Cambridge: MIT Press.
Kuhlmann, Marco, and Joakim Nivre. 2006. Mildly non-projective dependency structures. Paper presented at COLING/ACL 2006, Main Conference Poster Sessions, Sydney, Australia, July. pp. 507–14.
Kuperberg, Gina R., and T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31: 32–59. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2017. Investigating variation in island effects. *Natural Language and Linguistic Theory* 36: 1–37. [CrossRef] [PubMed]
Lakoff, George. 1986. Frame semantic control of the Coordinate Structure Constraint. In *Proceedings of the Chicago Linguistic Society 22*. Edited by Anne M. Farley, Peter T. Farley and Karl-Erik McCullough. Chicago: Chicago Linguistic Society. pp. 152–67.
Lasnik, Howard, and Tim Stowell. 1991. Weakest crossover. *Linguistic Inquiry* 22: 687–720.
Levine, Robert D. 2017. *Syntactic Analysis: An HPSG-Based Approach*. Cambridge: Cambridge Univeristy Press.
Levine, Robert D., and Thomas Hukari. 2006. *The Unity of Unbounded Dependency Constructions*. Stanford: CSLI Publications.
Levinson, Stephen C. 1987. Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics* 23: 379–434. [CrossRef]
Levinson, Stephen C. 1991. Pragmatic reduction of the binding conditions revisited. *Journal of Linguistics* 27: 107–61. [CrossRef]
Levy, Roger. 2005. Probabilistic Models of Word Order and Syntactic Discontinuity. Ph.D. thesis, Stanford University, Stanford, CA, USA.
Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106: 1126–77. [CrossRef]
Levy, Roger. 2013. Memory and surprisal in human sentence comprehension. In *Sentence Processing*. London: Psychology Press, pp. 78–114.
Levy, Roger, and T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems* 19: 849.
Levy, Roger P., and Frank Keller. 2013. Expectation and locality effects in german verb-final structures. *Journal of Memory and Language* 68: 199–222. [CrossRef]
Lewis, Richard. 1993. An architecturally-based theory of human sentence comprehension. Paper presented at 15th Annual Conference of the Cognitive Science Society, Boulder, CO, USA, June 18–21. Mahwah: Erlbaum, pp. 108–13.
Lewis, Richard. 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25: 93–115. [CrossRef]
Lewis, Richard, and Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29: 1–45. [CrossRef]
Lewis, Richard, Shravan Vasishth, and Julie Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10: 447–54. [CrossRef]
Liu, Haitao. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science* 9: 159–91.
Liu, Haitao, Chunshan Xu, and Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews* 21: 171–93. [CrossRef] [PubMed]
Michaelis, Laura A. 2012. Making the case for construction grammar. In *Sign-Based Construction Grammar*. Edited by Hans C. Boas and Ivan A. Sag. Stanford: CSLI Publications, pp. 31–68.
Miller, George, and Noam Chomsky. 1963. Finitary models of language users. In *Handbook of Mathematical Psychology*. Edited by R. Duncan Luce, Robert R. Bush and Eugene Galanter. New York: Wiley, vol. 2, pp. 419–92.
Müller, Gereon. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41: 35–82. [CrossRef]
Müller, Geron. 2018. Freezing in complex pre-fields. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Jäger, Andreas Konietzko and Susanne Winkler. Berlin: De Gruyter, pp. 105–39.
Nakamura, Michiko, and Edson T. Miyamoto. 2013. The object before subject bias and the processing of double-gap relative clauses in Japanese. *Language and Cognitive Processes* 28: 303–34. [CrossRef]
Newmeyer, Frederick J. 2016. Nonsyntactic explanations of island constraints. *Annual Review of Linguistics* 2: 187–210. [CrossRef]

Nunes, Jairo, and Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3: 20–43. [CrossRef]
Oda, Hiromune. 2017. Two types of the Coordinate Structure Constraint and rescue by PF deletion. *Proceedings of the North East Linguistic Society* 47: 343–56.
Oda, Hiromune. 2021. Decomposing and deducing the coordinate structure constraint. *The Linguistic Review* 38: 605–44. [CrossRef]
O'Grady, William, Miseon Lee, and Miho Choo. 2003. A subject-object asymmetry in the acquisition of relative clauses in korean as a second language. *Studies in Second Language Acquisition* 25: 433–48. [CrossRef]
Park, Kwonsik, Myung-Kwan Park, and Sanghoun Song. 2021. Deep learning can contrast the minimal pairs of syntactic data. *Linguistic Research* 38: 395–424.
Pesetsky, David. 1982. Paths and Categories. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In *The Representation of (In)definiteness*. Edited by Eric J. Reuland and Alice G. B. ter Meulen. Cambridge, MA: MIT Press, pp. 98–129.
Pesetsky, David. 1998. Some optimality principles of sentence pronunciation. In *Is the Best Good Enough?* Edited by Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky. Cambridge, MA: MIT Press, pp. 337–84.
Pesetsky, David. 2000. *Phrasal Movement and Its Kin*. Cambridge, MA: MIT Press.
Phillips, Colin. 2006. The real-time status of island phenomena. *Language* 82: 795–823. [CrossRef]
Phillips, Colin. 2013a. On the nature of island constraints i: Language processing and reductionist accounts. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Nobert Hornstein. Cambridge: University Press Cambridge, pp. 64–108.
Phillips, Colin. 2013b. Some arguments and non-arguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28: 156–87. [CrossRef]
Phillips, Colin, Matthew W. Wagers, and Ellen F. Lau. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. In *Experiments at the Interfaces*. Bingley: Emerald, vol. 37, pp. 147–80.
Pickering, Martin, and Guy Barry. 1991. Sentence processing without empty categories. *Language and Cognitive Processes* 6: 229–59. [CrossRef]
Pinker, Steven, and Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13: 707–27. [CrossRef]
Polinsky, Maria, Carlos G. Gallo, Peter Graff, Ekaterina Kravtchenko, Adam Milton Morgan, and Anne Sturgeon. 2013. Subject islands are different. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 286–309.
Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and CSLI Publications.
Postal, Paul. 1971. *Crossover Phenomena*. New York: Holt, Rinehart and Winston.
Pritchett, Bradley L. 1992. *Grammatical Competence and Parsing Performance*. Chicago: University of Chicago Press.
Pritchett, Bradley L. 1988. Garden path phenomena and the grammatical basis of language processing. *Language* 64: 539–76. [CrossRef]
Progovac, Ljiljana. 2016. A gradualist scenario for language evolution: Precise linguistic reconstruction of early human (and neandertal) grammars. *Frontiers in Psychology* 7: 1714. [CrossRef] [PubMed]
Pullum, Geoffrey K. 2019. What grammars are, or ought to be. Paper presented at 26th International Conference on Head-Driven Phrase Structure Grammar, București, Romania, July 25–26. Edited by Stefan Müller and Petya Osenova. Stanford: CSLI Publications, pp. 58–79.
Rajkumar, Rajakrishnan, Marten van Schijndel, Michael White, and William Schuler. 2016. Investigating locality effects and surprisal in written English syntactic choice phenomena. *Cognition* 155: 204–32. [CrossRef] [PubMed]
Reinhart, Tanya, and Eric Reuland. 1991. Anaphors and logophors: An argument structure perspective. In *Long-Distance Anaphora*. Edited by Jan Koster and Eric Reuland. Cambridge: Cambridge University Press, pp. 283–322.
Reis, Marga. 1993. *Wortstellung und Informationsstruktur*. Berlin: Walter de Gruyter.
Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.
Roberts, Craige. 2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5: 6–1. [CrossRef]
Rochemont, Michael. 1989. Topic islands and the subjacency parameter. *Canadian Journal of Linguistics—Revue Canadianne de Linguistique* 34: 145–70. [CrossRef]
Rochemont, Michael, and Peter W. Culicover. 1990. *English Focus Constructions and the Theory of Grammar*. Cambridge: Cambridge University Press.
Ross, John R. 1967. Constraints on Variables in Syntax. Ph.D. thesis, MIT, Cambridge, MA, USA.
Ross, John R. 1987. Islands and syntactic prototypes. *Chicago Linguistic Society Papers* 23: 309–20.
Sabel, Joachim. 2002. A minimalist analysis of syntactic islands. *Linguistic Review* 19: 271–315. [CrossRef]
Safir, Ken. 2017. Weak crossover. In *The Wiley Blackwell Companion to Syntax*. Edited by Martin Everaert and Henk van Riemsdijk. Hoboken, NJ: Wiley Online Library, pp. 1–40.
Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33: 431–84. [CrossRef]
Sag, Ivan A. 2012. Sign-based construction grammar—A synopsis. In *Sign-Based Construction Grammar*. Edited by Hans C. Boas and Ivan A. Sag. Stanford: CSLI Publications, pp. 61–197.
Sag, Ivan A., Inbal Arnon, Bruno Estigarribia, Philip Hofmeister, T. Florian Jaeger, Jeanette Pettibone, and Neal Snider. 2006. *Processing Accounts for Superiority Effects*. Stanford, CA: Stanford University, Unpublished ms.

Sag, Ivan A., and Janet D. Fodor. 1994. Extraction without traces. In *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics, Stanford University*. Edited by Raul Aranovich, Willian Byrne, Susanne Preuss and Martha Senturia. Stanford University: CSLI Publications. pp. 365–84.
Sag, Ivan A., Philip Hofmeister, and Neal Snider. 2007. Processing complexity in Subjacency violations: The complex noun phrase constraint. Paper presented at 43rd Annual Meeting of the Chicago Linguistic Society, Chicago, IL, May 3–5. Chicago: University of Chicago.
Sauerland, Uli. 1999. Erasability and interpretation. *Syntax* 2: 161–88. [CrossRef]
Selkirk, Elisabeth. 2011. The syntax-phonology interface. In *The Handbook of Phonological Theory*. Edited by John A. Goldsmith, Jason Riggle and Alan C. L. Yu. Oxford: Wiley-Blackwell, vol. 2, pp. 435–83.
Shain, Cory, Idan Asher Blank, Marten van Schijndel, William Schuler, and Evelina Fedorenko. 2020. fmri reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia* 138: 107307. [CrossRef] [PubMed]
Shan, Chung-Chieh, and Chris Barker. 2006. Explaining crossover and superiority as left-to-right evaluation. *Linguistics and Philosophy* 29: 91–134. [CrossRef]
Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1: 118–29. [CrossRef]
Sprouse, Jon, Matt Wagers, and Colin Phillips. 2012a. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88: 82–123. [CrossRef]
Sprouse, Jon, Matt Wagers, and Colin Phillips. 2012b. Working-memory capacity and island effects: A reminder of the issues and the facts. *Language* 88: 401–7. [CrossRef]
Staum Casasanto, Laura, Philip Hofmeister, and Ivan A. Sag. 2010. Understanding acceptability judgments: Distinguishing the effects of grammar and processing on acceptability judgments. Paper presented at 32nd Annual Conference of the Cognitive Science Society, Portland, OR, USA, August 11–14. Edited by Stellan Ohlsson and Richard Catrambone. Austin: Cognitive Science Society, pp. 224–29.
Szabolcsi, Anna, and Terje Lohndal. 2017. Strong vs. weak islands. In *The Wiley Blackwell Companion to Syntax*, 2nd ed. Edited by Martin Everaert and Henk van Riemsdijk. New York: John Wiley.
Temperley, David. 2007. Minimization of dependency length in written english. *Cognition* 105: 300–33. [CrossRef]
Trotzke, Andreas, Markus Bader, and Lyn Frazier. 2013. Third factors and the performance interface in language design. *Biolinguistics* 7: 1–34. [CrossRef]
Truckenbrodt, Hubert. 1995. Phonological Phrases: Their Relation to Syntax, Focus, and Prominence. Ph.D. thesis, MIT, Cambridge, MA, USA.
Ueno, Mieko, and Susan M. Garnsey. 2008. An ERP study of the processing of subject and object relative clauses in japanese. *Language and Cognitive Processes* 23: 646–88. [CrossRef]
van Dyke, Julie, and Richard Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49: 285–316. [CrossRef]
van Schijndel, Marten, Andy Exley, and William Schuler. 2013. A model of language processing as hierarchic sequential prediction. *Topics in Cognitive Science* 5: 522–40. [CrossRef] [PubMed]
Varaschin, Giuseppe. 2021. A Simpler Syntax of Anaphora. Ph.D. thesis, Universidade Federal de Santa Catarina, Florianopolis, Brazil.
Varaschin, Giuseppe, Peter W. Culicover, and Susanne Winkler. 2022. In pursuit of condition C. In *Information Structure and Discourse in Generative Grammar: Mechanisms and Processes*. Edited by Andreas Konietzko and Susanne Winkler. Berlin: Walter de Gruyter, *to appear*.
Vasishth, Shravan, and Richard L. Lewis. 2006. Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82: 767–94. [CrossRef]
Vasishth, Shravan, Bruno Nicenboim, Felix Engelmann, and Frank Burchert. 2019. Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences* 23: 968–82. [CrossRef] [PubMed]
Villata, Sandra, Luigi Rizzi, and Julie Franck. 2016. Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua* 179: 76–96. [CrossRef]
Warren, Tessa, and Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85: 79–112. [CrossRef]
Wasow, Thomas. 1979. *Anaphora in Generative Grammar*. Amsterdam: John Benjamins Publishing Company.
Wexler, Kenneth, and Peter W. Culicover. 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.
Winkler, Susanne, Janina Radó, and Marian Gutscher. 2016. What determines 'freezing' effects in *was-für* split constructions? In *Firm Foundations: Quantitative Approaches to Grammar and Grammatical Change*. Edited by Sam Featherston and Yannick Versley. Boston and Berlin: Walter de Gruyter, pp. 207–31.
Yadav, Himanshu, Samar Husain, and Richard Futrell. 2021. Do dependency lengths explain constraints on crossing dependencies? *Linguistics Vanguard* 7: 1–7. [CrossRef]
Zwicky, Arnold M. 1986. The unaccented pronoun constraint in English. *Ohio State University Working Papers in Linguistics* 32: 100–13.

languages

MDPI

Article

Extraction from Present Participle Adjuncts: The Relevance of the Corresponding Declaratives

Andreas Kehl

English Department, Universität Tübingen, Wilhelmstraße 50, 72074 Tübingen, Germany; andreas.kehl@uni-tuebingen.de

Abstract: In this article, I will argue that many of the theoretical approaches to extraction from participle adjunct islands suffer from the fact that the focus of investigation lies on perceived grammaticality differences in interrogative structures. Following approaches which make an explicit connection between extraction asymmetries and properties of the underlying proposition, I will argue that there is good evidence for the existence of similar differences in declarative adjunct constructions which can explain most of the grammaticality patterns observed for interrogatives. A crucial distinction to the majority of previous theories is the focus on acceptability rather than grammaticality, and the assumption that acceptability in declaratives is determined by a variety of semantic and syntactic complexity factors which do not influence how strongly extraction degrades the structure. This line of argumentation is more compatible with approaches to island phenomena that explain the low acceptability of some extractions by independent effects such as processing complexity and discourse function instead of syntactic principles blocking the extraction. I will also discuss a partially weighted, multifactorial model for the acceptability of declarative and interrogative participle adjunct constructions, which explains the judgment patterns in the literature without the need for additional, complex licensing conditions for extraction.

Keywords: adjunct islands; *wh*-extraction; locality; present participle; gradient acceptability; acceptability model

Citation: Kehl, Andreas. 2022. Extraction from Present Participle Adjuncts: The Relevance of the Corresponding Declaratives. *Languages* 7: 177. https://doi.org/10.3390/languages7030177

Academic Editors: Anne Mette Nyvad and Ken Ramshøj Christensen

Received: 22 December 2021
Accepted: 15 June 2022
Published: 8 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Since the formulation of the Condition on Extraction Domain (CED, Huang 1982), more and more apparent counterexamples to this strict locality condition have surfaced, including extractions from subjects and adjuncts that are judged as grammatical. Compare the ungrammatical extraction from an adverbial clause in (1a) with the extraction from an adjunct headed by a present participle in (1b), which is considered grammatical; the (participial) adjunct predicate is shown in square brackets in most of the examples used in this article. An acceptable extraction from subject is shown in (2).

(1) a. *Who did Mary cry [after John hit t]? (Huang 1982, p. 503)
b. What$_i$ did John arrive [whistling t$_i$]? (Borgonovo and Neeleman 2000, p. 200)

(2) What did [$_{IP}$ [$_{NP}$ the attempt to find __] end in failure]? (Hofmeister and Sag 2010, p. 370)

Attested examples of extraction from participle adjuncts, as in (1b), are often found in the form of relativization, as in (3) from Santorini (2019) and (4) from a news article; in these cases, it is a nominal element that is associated with a gap site in the complement position of a participle adjunct instead of a *wh*-pronoun.

(3) a. The magazine I spend most of my days [reading __].

b. "This is the game I grew up [watching __]," Wilson added.

(Santorini 2019)

(4) Already uncomfortable with the policy—the USD 1.3 trillion legislation includes massive spending hikes that contradict prior GOP complaints about the debt—many Republicans were left dumbfounded by [a process]$_i$ that looked a lot like one they had won office [criticizing $\_\_{}_i$]. (Washington Post, 22 March 2018)

I follow Truswell (2011, p. 30) in referring to adjunct constructions, such as (1b), (3), and (4) as Bare Present Participle Adjuncts (BPPA); they are characterized by an untensed present participle as the head of the adjunct predicate, as well as the absence of an explicitly encoded subject or subordinators.

However, not all BPPA constructions allow extraction as easily as (1b): minimally different examples, such as (5), are reported to block the extraction.

(5) *What$_i$ did John dance [imagining t$_i$]? (Borgonovo and Neeleman 2000, p. 199)

There are, thus, two problems to be addressed: (i) apparently grammatical extractions from adjuncts that should be excluded by the CED (1a vs. 1b), and (ii) variation within an adjunct type where some extractions are allowed while others are not (1b vs. 5). The first problem has been addressed in the Minimalist literature in two ways: first, abandoning or modifying the original formulation of the CED to accommodate such cases,[1] and second, to reconsider the adjunct status of apparent counterexamples to the CED (e.g., Graf 2015). These approaches still assume that there is a syntactic principle at work which determines when extraction is possible. A major alternative to this syntactic perspective is taken by approaches which do not assume a syntactic principle behind extraction asymmetries, but rather more general principles. This includes approaches based on processing (e.g., Sag et al. 2008; Hofmeister and Sag 2010), information structure (Goldberg 2006, 2013), pragmatics (Chaves and Putnam 2020), and discourse functions (Abeillé et al. 2020). Such approaches line up with the Radical Unacceptability Hypothesis proposed in Culicover et al. (2022), to which I return at the end of this article.

I will focus on the second problem in the rest of this article.[2] For reasons of space, the discussion focuses on examples with *wh*-extraction, but it should be kept in mind that there is growing evidence that different types of filler–gap dependencies yield different effects, so that so-called island constraints do not appear to be cross-constructionally active (Liu et al. 2022); see also Sprouse et al. (2016) and Abeillé et al. (2020) for findings and discussion, as well as Kehl (2021, experiment 1) for a comparison between declarative, interrogative, and relativized BPPA constructions. Differences between types of extractions become all the more relevant since much of the existing literature focuses on *wh*-extraction, whereas many attested examples are instances of relativization (Chaves and Putnam 2020; Santorini 2019). I will briefly address other dependency types at the end of Section 5.

The variation in the extraction behavior of interrogative BPPA constructions has resulted in several approaches that try to find an explanation for such patterns; the influential theoretical approaches in Borgonovo and Neeleman (2000) and Truswell (2007, 2011) propose licensing conditions to accommodate this island-internal variation. In this article, I will follow the discussion in Brown (2017) and Kehl (2021), agruing that the *grammaticality* patterns observed for extraction from BPPA constructions are actually a reflection of different degrees of *acceptability* which are already observable in the declarative counterparts and evoke the impression of grammaticality differences in interrogatives. Thus, I assume that the acceptability difference between (1b) and (5) is equal to that between the declaratives in (6):

(6) a. John arrived [whistling the Blue Danube].

b. John danced [imagining the Gobi Desert].

(Borgonovo and Neeleman 2000, pp. 199–200)

By extension, I also assume that the same acceptability contrast between verbs such as *arrive* and *dance* is visible in other dependency types, such as the relative clauses in (7), which are more similar to the attested data in (3) and (4) above:

(7) a. This is [the song]$_i$ that John arrived [whistling $\_\__i$].

b. This is [something]$_i$ that John danced [imagining $\_\__i$].

The basic idea behind this assumption, which I will argue for in the remainder of this article, is that differences between the two main verbs *arrive* and *work* result in different degrees of acceptability independently of whether the sentence form is declarative, relativization, or interrogative. In other words, the relative acceptability of the declaratives are good predictors for relative acceptability in different sentence forms; see also Chaves and King (2019), who find a relation between judgments of relevance and acceptability of subextraction from subjects. This line of research shifts the focus of attention to the semantic and/or pragmatic factors which affect acceptability in the underlying declarative structures. This comparison of extraction from island constructions to possible differences in the underlying declaratives ties into the growing body of research that does not focus on extraction constructions alone (among others, Abeillé et al. 2020; Brown 2017; Chaves and King 2019; Chaves and Putnam 2020). The relevance of drawing on more subtle differences in declaratives to explain differences at the fringes of grammaticality in extraction structures goes back to at least Kuno (1987), an idea that is picked up prominently in the pragmatic approach to extraction asymmetries in Chaves and Putnam (2020), but also the discussion of complexity differences in Culicover and Winkler (2022).

The discussion in this article centers around the question whether extraction asymmetries observed for BPPAs need to be captured by a grammatical principle, as proposed in Borgonovo and Neeleman (2000) and Truswell (2007, 2011), or whether these asymmetries can be explained independently. I argue for the second position and discuss the possibility of capturing judgment differences, such as (1b) vs. (5); the underlying idea is that the semantic compatibility between the two predicates in this construction affects acceptability both in the presence and absence of a dependency such as *wh*-extractionextraction. The discussion of this narrow set of examples is closely related to the more general proposal in Culicover and Winkler (2018, 2022) and Culicover et al. (2022) that many instances of such judgment differences in extraction phenomena can be accounted for without the need to introduce grammatical principles.

This article is structured as follows: I will first provide a short summary of the grammaticality patterns reported in Borgonovo and Neeleman (2000) as well as Truswell (2007) in Section 2 as a basis for the remainder of the discussion; in Section 3, I discuss the relations between the concepts of grammaticality and acceptability, as well as the potential mapping problems between gradient and binary judgments; I will then suggest a factorial design for the detection of island-internal variation that allows for an experimental validation of factors that are assumed to influence how strong extraction affects different types of declaratives; Section 4 examines previous experimental studies which compare declarative and interrogative BPPA constructions and whether their results speak for or against the conclusions in the theoretical literature; Section 5 then discusses factors which influence the acceptability of declarative BPPA constructions independently of extraction and combine these factors into an acceptability model for declarative and interrogative BPPA constructions; in Section 6, I take a brief look at evidence from related phenomena that also come to the conclusion that differences in declaratives have an impact on theory development; Section 7 concludes this article.

2. Reported Grammaticality Patterns

In this section, I will summarize the reported grammaticality patterns for extraction from participle adjuncts in two influential accounts: Borgonovo and Neeleman (2000) and Truswell (2007).[3] Both accounts share the intuition that different grammaticality patterns exist in interrogatives which are not present in declaratives; this leads them to propose additional licensing mechanisms for extraction to accommodate these interrogative patterns. I will not go into the technical details of these accounts for reasons of space and because the focus of this article is on the relation between declaratives and interrogatives instead of the licensing mechanisms they propose. As I will show in Section 3, such a comparison uncovers problematic aspects of these accounts.

2.1. Transparency Depends on Verb Types

Borgonovo and Neeleman (2000) report on a grammaticality pattern that allows extraction from participial adjuncts modifying unaccusative and reflexive transitive main verbs, as in (8a) and (8b); in contrast, extraction from adjuncts modifying unergative and non-reflexive transitives, as in (8c) and (8d), results in ungrammaticality.

(8) a. What$_i$ did John arrive [whistling t$_i$]? [unaccusative]

b. What$_i$ did John hurt himself [trying to fix t$_i$]? [reflexive transitive]

c. *What$_i$ did John dance [imagining t$_i$]? [unergative]

d. *What$_i$ did John hurt Bill [trying to fix t$_i$]? [non-reflexive transitive]

(Borgonovo and Neeleman 2000, pp. 199–200)

The main proposal resulting from this pattern is that some verb types are able to L-mark adjuncts by means of a syntactic reflexivity relation where the internal argument DP binds both the θ-roles of the adjunct predicate and the main verb. The adjunct will then count as L-marked and obeys the CED because it is properly governed. Only unaccusatives and reflexive transitives are able to L-mark the adjunct because the right structural configuration is only possible with an internal argument that is also the external argument of the adjunct predicate. Unergatives fail to L-mark the adjunct because they do not have an internal argument and do not project the necessary V′-layer (Borgonovo and Neeleman 2000, pp. 212–13); L-marking is not possible for non-reflexive transitives because the external argument of the adjunct is also the external argument of the main verb. In both cases, extraction is banned by the CED because the adjunct is not L-marked.

Crucially, L-marking is a condition that is specific to the licensing of extraction: it does not have an effect in declaratives because it is irrelevant there. Therefore, the declarative sentences in (9) underlying the interrogatives in (8c) and (8d) are completely unmarked.

(9) a. John danced [imagining the Gobi Desert].

b. John hurt Bill [trying to fix the roof].

(Borgonovo and Neeleman 2000, pp. 199–200)

Because declarative BPPA constructions are unconstrained in terms of grammaticality differences, the source of ungrammaticality in interrogatives is caused by the extraction operation itself, which fails to be licensed if L-marking cannot be established for unergatives and non-reflexive transitives. The required adjustments to subjacency-based locality theory are modest and can be expressed in core-syntactic terms, even if the theory requires ternary branching to establish syntactic reflexivity between the verb, its internal argument, and the adjunct predicate. Still, a major problem with this account is that it does not consider any potential variation in the declarative counterparts and exclusively relies on extraction-related factors to explain the pattern in interrogative BPPA constructions.

2.2. Transparency Depends on Telicity

A slightly different pattern is described in Truswell (2007), who focuses on the event structure of the BPPA construction. The key proposal is that extraction from an adjunct predicate is only licensed in the grammar if the adjunct fills an open or underspecified event position in the event structure of the matrix predicate. This means two things: (i) the matrix predicate needs to encode at least two subevents, and (ii) one of these is underspecified by the lexical semantics of the matrix predicate. The two event types that encode more than one subevent are achievements and accomplishments in terms of Vendler (1957); they are composed of a culmination point and a durative subevent leading up to this endpoint, which is optional for achievements; see Rothstein (2004). States and activities, on the other hand, either encode no event at all (states) or only a single subevent (activities). In case the adjunct can be interpreted as supplying more information about the underspecified subevent, the two predicates describe facets of a single event, mirroring the lexical semantics of a maximally complex verb (Truswell 2007, p. 1369). This amounts to the generalization that extraction from the adjunct is only possible if the matrix predicate is telic; this derives the predictions for the contrast in (10) with the atelic verb *work* (10a) and the telic *arrive* (10b):

(10) a. *What did John work whistling __ ? [atelic matrix predicate]

b. What did John arrive whistling __ ? [telic matrix predicate]

(see Truswell 2007, p. 1369)

These predictions are similar to those in Borgonovo and Neeleman (2000), but formulated in event-semantic terms, which are not exclusively tied to extraction. In addition to achievement matrix predicates, such as (10b), extraction is also possible from structures with accomplishment main verbs, such as in (11), provided that the adjunct can describe the cause of the matrix predicate:

(11) What did John drive Mary crazy [trying to fix t]? (Truswell 2007, p. 1356)

Like Borgonovo and Neeleman (2000), Truswell (2007) concludes that the corresponding declaratives in (12) do not show a similar pattern and that the grammaticality pattern in interrogatives is the result of extraction. Both accounts do consider declarative counterparts with respect to their grammaticality, but do not observe significant differences in acceptability.[4]

(12) a. John worked [whistling a song].

b. John arrived [whistling a song].

(see Truswell 2007, pp. 1369, 1373)

This means that the syntactic extraction operation needs to be sensitive to the distinctions between different event types, but also to the lexical semantics of the two predicates, as well as potential causal chains between them. Unless information about the aspectual type and causality are directly encoded syntactically, as, for example, in Borer (2005) and Ramchand (2008), this extraction pattern is impossible to explain in core syntactic terms. It is not an immediate problem that Truswell (2007) considers both sentences in (12) grammatical, but this focus on grammaticality requires the formulation of extraction conditions in event-semantic terms (or a post-syntactic event-semantic output filter, as suggested in Truswell 2011).

Both accounts sketched in this section agree that declarative BPPA constructions are relatively unconstrained with respect to grammaticality differences and that the pattern in interrogatives is a direct result of failures in the licensing mechanism for extraction. I will argue in the following section that this perspective overestimates the reported grammaticality differences in interrogatives, and at the same time underestimates potential differences

in the declarative counterparts. The main reason for these problematic aspects is rooted in the distinction between the concepts of grammaticality and acceptability, as well as the relation between gradient and binary judgments.

3. Grammaticality, Acceptability, and the Relation between Declaratives and Interrogatives

In this section, I will discuss problematic aspects of the exclusive focus on *grammaticality* differences in interrogatives without also considering potential *acceptability* differences in the declarative counterparts. The problem is one of mapping relations between binary grammaticality judgments and gradient acceptability judgments, because sentences that receive the same binary grammaticality marking may still show significant differences in acceptability that are not properly represented in all grammaticality judgments. For example, it is reasonable to consider both examples in (13) grammatical, but experimental evidence suggests that (13a) is less acceptable than (13b). Among others, the lower acceptability and negative impact on online sentence processing of additional arguments is shown in Jurka (2010, 2013), Polinsky et al. (2013), Brown (2017), and Culicover and Winkler (2022). An additional issue in (13a) is that there is a degree of ambiguity whether the adjunct refers to John or Bill. In connection with syntactic dependencies, the greater processing cost and, thus, reduced acceptability is predicted by Dependency Locality Theory (Gibson 1998, 2000); see also Section 4.[5]

(13) a. John hurt Bill [trying to fix the roof].

b. John arrived [whistling the Blue Danube].

(Borgonovo and Neeleman 2000, p. 200)

Subsection 3.1 describes the contrast between binary grammaticality and gradient acceptability judgments, as well as their relation in more detail; the focus here is on which conclusions can be drawn from these two measurements and the risk of not distinguishing between them properly. Subsection 3.2 proposes an adapted factorial experiment design that allows for the investigation of island-internal variation which includes a comparison to the declarative base position. Subsection 3.3 emphasizes the usefulness of including standardized reference fillers in acceptability judgment tasks for conceptual and methodological reasons.

3.1. Gradient and Binary Judgments

One of the core issues in the evaluation of the theoretical approaches in Borgonovo and Neeleman (2000) and Truswell (2007, 2011) lies in the distinction between the concepts of *grammaticality* and *acceptability* discussed in Chomsky (1965). Chomsky (1965) models this distinction as one between *competence* and *performance*: the former refers to those aspects of language that are part of a speaker's grammar, whereas the latter reflects the use of language that is also affected by other factors. Grammaticality is seen as a measure of whether a sentence is licensed by the grammar; this evaluation has often been considered to be a categorical distinction, even though Chomsky (1965, p. 11) already notes that it is probably "a matter of degree". Acceptability as a measure of naturalness and comprehensibility does not solely depend on grammaticality, but grammaticality is one of the factors that determine acceptability: a sentence that is considered grammatical can still show low acceptability because they are semantically or pragmatically anomalous, or because they are difficult to process (Chomsky 1965, p. 11). Ungrammaticality refers to the fact that a given structure cannot be computed by the grammar, or runs afoul at the interfaces, for example because not all uninterpretable features are checked and deleted during the derivation. Acceptability is partially fed by grammaticality, but also affected by additional factors that are independent of grammaticality: as is well known, there are sentences which can be generated by the grammar but can be anomalous semantically and/or pragmatically, or pose processing difficulties that impact acceptability judgments. On the other hand,

there are sentences which are grammatically ill-formed but appear intuitively acceptable, a phenomenon called 'illusions of grammaticality' in Phillips (2013, p. 106).

There is, thus, a mapping problem between grammaticality and acceptability because not all sentences that are considered grammatical are necessarily equally acceptable, also noted in Chomsky (1965, p. 11). Especially problematic are cases where acceptability is on the borderline or threshold of grammaticality: minimally different acceptable sentences run the risk of being assigned opposite grammaticality judgments, even if the relative distance in acceptability between them is smaller than the distance between two fully grammatical or ungrammatical sentences. I will elaborate on this problem in the remainder of this subsection.

Consider the two declarative BPPA constructions in (14), with an unaccusative (14a) and an unergative (14b) matrix predicate. The predictions of Truswell (2007) and Borgonovo and Neeleman (2000) agree on the fact that extraction from the adjunct in (14a) will be grammatical, whereas extraction from (14b) will not.

(14) a. John arrived whistling a funny song.
b. John worked whistling a funny song.

Let us assume a gradient Likert-type judgment scale with seven discrete points, and a binary categorization into grammatical and ungrammatical sentences. Let us also assume that gradient judgments on or above the middle of the gradient scale, i.e., every gradient judgment ≥ 4, will be mapped to the binary judgment 'grammatical', and that gradient judgments < 4 will be mapped to 'ungrammatical'. Thus, if (14a) is assigned a gradient judgment of 7 and (14b) a judgment of 5, both will be mapped onto a grammatical binary judgment; this is shown in (15).

(15) a. (14a) $\rightarrow$ gradient judgment: 7 $\rightarrow$ binary judgment: grammatical
b. (14b) $\rightarrow$ gradient judgment: 5 $\rightarrow$ binary judgment: grammatical

This is, in essence, what Borgonovo and Neeleman (2000) and Truswell (2007, 2011) assume about declarative BPPA constructions, with a focus on the outcome of the binary grammaticality judgment. For now, it is not immediately relevant why (14b) should be less acceptable on a gradient scale compared to (14a). The data reported in Brown (2017) and Kehl (2021) support the assumption that there is a statistically significant acceptability difference between the two, even if this difference might not be as pronounced as in this hypothetical example.

When extraction takes place from the adjunct, the gradient judgment will decrease for both structures because the formation and resolution of filler–gap dependencies is a cognitively costly operation and because interrogatives are semantically more complex than declaratives (Chaves and Putnam 2020; Hofmeister and Sag 2010; Wagers 2013). Since the extraction domain is an adjunct, this judgment decrease will probably be larger compared to extraction from a subcategorized complement, as predicted by the CED.[6] The interrogative counterparts of (15) are shown in (16), without judgment marks.

(16) a. What did John arrive whistling?
b. What did John work whistling?

A final assumption made here, again supported by the experimental evidence in Brown (2017) and Kehl (2021), is that both structures are affected to the same degree by extraction, meaning that the decrease in the gradient judgment is identical; the gradient judgment for (16b) will then fall below the threshold in the middle of the scale, resulting in an ungrammatical binary judgment. For (16a), the gradient judgment remains on or above the threshold, yielding a grammatical binary judgment; this is shown schematically in (17).

(17) a. (16a) → gradient judgment: 5 → binary judgment: grammatical

b. (16b) → gradient judgment: 3 → binary judgment: *ungrammatical

On the surface, this results in exactly the grammaticality patterns constructed in Borgonovo and Neeleman (2000) and Truswell (2007). However, what we are mostly interested in is whether the relative differences in gradient judgments between the two sentence pairs are identical or whether one is larger than the other. To check for this, we subtract the two declarative judgments from one another and compare this to the same difference between the interrogative counterparts. If the difference pairs are equal to each other (or at least not significantly different), then there is no need for additional licensing mechanisms for extraction because the gradient judgment differences in interrogatives can be predicted from the differences in declaratives in a linear additive way. This is shown in (18a) and is the simpler case because then the only explanation required is what causes the differences in declaratives plus the independent decrease caused by extraction. If, on the other hand, the relative differences are of different magnitudes, as shown in (18b) and (18c), then this requires an explanation for this additional difference that cannot be predicted from the gradient contrasts in declaratives. These patterns can be referred to as superadditive and subadditive. Depending on whether the difference between declaratives is smaller than that for interrogatives, as in (18b), or the other way round, as in (18c), this leads to the need either for additional licensing mechanisms or a repair mechanism, respectively.

(18) a. $\frac{\text{differences between declaratives}}{\text{differences between interrogatives}} = 1$ [no licensing mechanism required]

b. $\frac{\text{differences between declaratives}}{\text{differences between interrogatives}} < 1$ [licensing mechanism required]

c. $\frac{\text{differences between declaratives}}{\text{differences between interrogatives}} > 1$ [repair mechanism required]

This metric is similar to the differences-in-differences score employed in Sprouse et al. (2012, 2013), which isolates the effect sizes of individual factors and evaluates whether the combination of two factors negatively impact acceptability scores to a greater (or lesser) degree than the two individual factors.

Figure 1 illustrates the first two possibilities in (18): the left panel corresponds to (18a) where the gradient judgment differences are identical for both types of matrix predicate; the right panel shows the pattern where the decrease caused by extraction is larger for atelic matrix predicates than that for telic predicates (18b). I omit the case of (18c) for expository purposes. The shaded area in Figure 1 shows the range of the gradient scale that will be mapped onto an ungrammatical binary judgment. The experimental results in both Brown (2017) and Kehl (2021, experiment 2) correspond more closely to the pattern on the left rather than the one on the right, showing that the strength of the acceptability decrease in interrogatives is not influenced by the other factors they investigate. I will return to this discussion in Section 4 below.

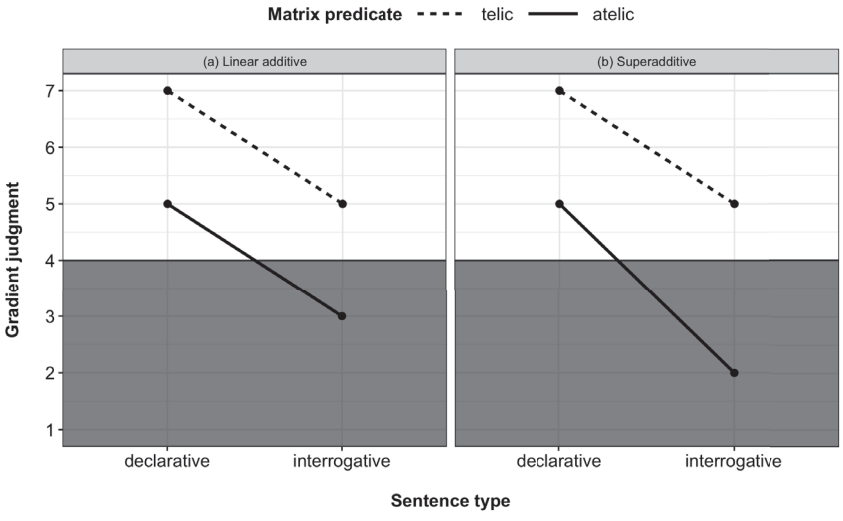


Figure 1. Schematic illustration of the linear additive and superadditive gradient judgment patterns in (18); the shaded area shows the part of the scale that will be mapped to ungrammatical judgments.

The difference between these patterns is obscured if the sole focus is on the binary judgment because this ignores potential differences in gradient judgments for declaratives; in a sense, some information is lost in the mapping between gradient and binary judgments. On its own, this is not problematic, but it becomes so if used as a basis for the postulation of licensing mechanisms for extraction. One possibility to avoid the potential pitfalls of binary judgments is to broaden the data pool and gather binary judgments from multiple informants, which can then be converted to a gradient scale similar to Likert-scales by calculating the ratio of grammatical-to-ungrammatical responses (Bader and Häussler 2010, 2019). This method has been shown to result in similar patterns as judgments on discrete or continuous scales.

Taking a step back, the binary judgment differences in Borgonovo and Neeleman (2000) and Truswell (2007) can be converted to acceptability measures, meaning that the grammatical extractions are more acceptable than the ungrammatical ones. However, the formulation as grammaticality judgments runs the risk of leading to proposals about the architecture of the syntactic component and its interfaces with semantics and pragmatics. Therefore, I think that it is advisable to focus on acceptability first and then reason about the model of grammar that best fits with these results.

3.2. A Factorial Design for Island-Internal Variation

The procedure described in the previous section represents a modification of the factorial design for island effects in Sprouse et al. (2012, 2013) and Kush et al. (2018, 2019). The original design compares conditions in a way that allows to isolate the individual effects of two factors: the difference between extraction from matrix clauses vs. embedded domains and between extraction from non-island vs. island domains. See (19) for an illustration of the design:

(19) a. Who __ thinks that John bought a car? [NON-ISLAND/MATRIX]
b. What do you think that John bought __ ? [NON-ISLAND/EMBEDDED]
c. Who __ wonders whether John bought a car? [ISLAND/MATRIX]
d. *What do you wonder whether John bought __ ? [ISLAND/EMBEDDED]
(Sprouse et al. 2013, p. 25)

This design allows quantifying three sets of contrasts and the respective effects they have on acceptability: (i) the contrast between (19a)–(19b) isolates a possible effect between extraction from the matrix clause vs. the embedded clause; (ii) the contrast between (19a)–(19c) detects whether the presence or absence of an island domain, in this case a *wh*-island introduced by *whether*, affects acceptability; and (iii) the contrast between (19b)–(19d) compares the cost of extraction from a non-island vs. from an island domain (see Sprouse et al. 2013, p. 25). Often, theoretical approaches will focus on the contrast between (19b) vs. (19d) and conclude that *whether*-clauses are islands if this extraction feels less acceptable than the non-island. However, this leaves unaccounted the potential effect that the presence of a *whether*-clause has on acceptability independently of extraction.

To solve this, Sprouse et al. (2012, 2013) include this effect in the calculation of potential island effects: if the acceptability judgment for the 'worst' condition (19d) compared to the unmarked reference condition (19a) cannot be predicted from the differences between (19a) and (19b) and (19a)–(19c), then this additional acceptability decrease is called an 'island effect' which needs to be accounted for theoretically.

The same reasoning can be applied to investigate the validity of theoretical approaches such as those in Borgonovo and Neeleman (2000) or Truswell (2007): instead of comparing an island construction with a non-island, two instances of the same island type are tested in declarative and interrogative conditions. They differ minimally in one of the factors isolated in the literature, such as event structure or the verb type of the matrix predicate. This allows to examine whether such factors determine how strongly extraction degrades acceptability, as well as whether there are acceptability differences in the declaratives that are the source of the reported differences in interrogatives. An example of such a design is given in (20), based on the example sentences discussed in the previous section. This relatively simple 2×2 design manipulates the matrix predicate as telic or atelic, as well as the difference between declarative and interrogative sentences. The manipulation of other factors, also ones with more levels, is of course also possible; for a more complex $2 \times 2 \times 2$ design that crosses the factors TELICITY, TRANSITIVITY, and EXTRACTION, see Brown (2017). For example, the simple comparison between declarative and interrogative sentence forms can be augmented to also include relative clauses and topicalization.

(20) a. John arrived whistling a funny song. [telic/−wh]
b. What did John arrive whistling? [telic/+wh]
c. John worked whistling a funny song. [atelic/−wh]
d. What did John work whistling? [atelic/+wh]

The statistical analysis will then compare the effects of the two factors, in this case telicity and extraction, as well as the interaction between them. The absence of a significant interaction indicates that the strength of extraction is not influenced by the factor that distinguishes the declarative conditions. If there is a significant interaction, additional licensing or repair mechanisms are called for, as explained in the previous section. Like the detection of island effects in the original factorial design in Sprouse et al. (2012, 2013), the question whether extraction from a 'suboptimal' adjunct island configuration leads to drops in acceptability that cannot be explained independently of extraction would lead to additional licensing requirements. Determining this need for licensing mechanisms should be at the core of investigations into island-internal variation and should be backed up with experimental data in addition to initial, intuitive judgments.

3.3. The Use of Standardized Fillers

The results of gradient judgment studies can sometimes be difficult to interpret. Typically, the experimental conditions are compared to each other in terms of significant differences between conditions in the data pool, or in terms of effect structures in the case of factorial designs. Although this is the main interest of an experimental study, i.e., to test hypotheses about acceptability contrasts and the influence of specific factors, it is also

of interest to compare where the experimental conditions are located on the continuum of gradient acceptability, regardless of whether this continuum is expressed in discrete Likert-type scales or truly continuous judgments as in Magnitude Estimation (Bard et al. 1996) or Thermometer judgments (Featherston 2020). One possibility is to add control conditions that are closely related to the construction under investigation, as implemented in Abeillé et al. (2020) with grammatical and ungrammatical controls.[7] In her experiment on extraction from adjuncts in English that is closely related to the design in (20), Brown (2017) includes grammatical and ungrammatical controls as in (21); as extraction from tensed adjuncts as in (21b) is not always considered unacceptable, extraction from a conjunct as in (22a) can also be used for unacceptable controls because there is general agreement in the literature that such extractions are ungrammatical (Liu et al. 2022). A resumptive pronoun at the gap site, as shown in (22b), can also be used to construct ungrammatical control conditions that are close to the design implemented (Chaves and Putnam 2020, pp. 218–19).

(21) a. Which ice cream did Mary eat before she saw the celebrity?
[grammatical control]

b. *Which celebrity did Mary eat an ice cream before she saw?
[ungrammatical control]
(Brown 2017, p. 120)

(22) a. *What did Mary go to work and whistle?

b. *What did Mary arrive at the office whistling it?

The set of standardized reference fillers developed for English in Gerbrich et al. (2019) are designed to provide anchor points along gradient or discrete judgment scales, ranging from a high level of acceptability to a low level; the idea of providing a standardized scale for acceptability is also found in Featherston (2009), who develops a set of German reference fillers.

The goal of the standardized fillers is to provide anchor points on the extremes of the rating scale with highly acceptable and highly degraded sentences, as well as a range of acceptability in between; ideally, this results in a reference scale with equal distances between the individual levels, so that the experimental items can be assigned a relative level of normed acceptability. The choice of very general levels of well-formedness along the spectrum of acceptability which is not limited to control items that are related to the construction has the advantage that the fillers can be re-used across multiple experiments and, thus, allows a more grounded discussion of acceptability across experiments. It is of course possible to include both the standard fillers and construction-specific control conditions in an experiment. A sample of the reference fillers is given in (23); the assignment of more traditional graded grammaticality judgment marks, such as '?' or '*', are adapted from Gerbrich et al. (2019, p. 310).

(23) a. √√A: The patient fooled the dentist by pretending to be in pain.

b. √B: Before every lesson the teacher must prepare their materials.

c. ?C: Hannah hates but Linda loves eating popcorn in the cinema.

d. ??D: Who did he whisper that had unfairly condemned the prisoner?

e. *E: Historians wondering what cause is disappear civilization.
(see Gerbrich et al. 2019, p. 315)

The two best levels A and B are usually not marked in such judgment schemes, and are both considered fully grammatical; still, Gerbrich et al. (2019) suggest that there are significant acceptability differences between these grammatical levels, which are difficult to detect in judgments with limited conventionalized markings. Judging from their experimental results with the standardized fillers, Gerbrich et al. (2019, p. 309) conclude that

there may be even more distinguishable levels of well-formedness. Note that the E-level is still interpretable, but highly unnatural; it is possible to add a further level with low interpretability, as for example in the adaptation of the standard fillers in Brown et al. (2021). Brown et al. (2021, p. 10) refer to this as a "clearly ungrammatical level" with examplessuch as *The ink was for spilled* that are considered both unacceptable and uninterpretable.

Figure 2 illustrates the expected distribution of the five sets of standardized fillers on a 7-point scale of acceptability; see also the discussion in Featherston (2020, pp. 168–72) showing a similar distribution in *z*-scores based on an actual experiment. The exact values may vary from experiment to experiment, and it may not always be the case that the distance between the levels is evenly distributed, especially if target conditions fall between two of the levels (Gerbrich et al. 2019, pp. 315–16). From these predicted values and the judgment marks in Gerbrich et al. (2019), it becomes apparent that the binary ungrammaticality marking may be limited to a rather small gradient acceptability area, unlike the assumption above that the threshold for binary grammaticality judgments lies in the middle of the gradient scale. I leave this point open for discussion here.

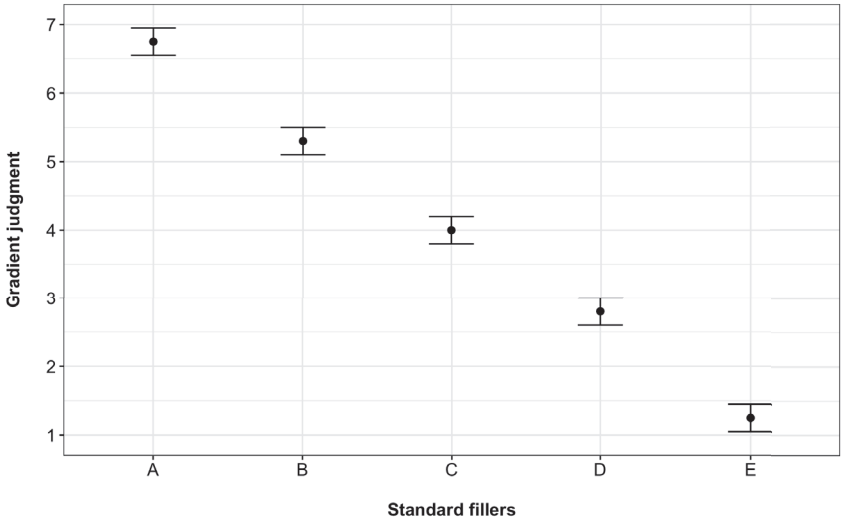


Figure 2. Expected rating distribution for the A–E levels of the standard fillers in (23) on a gradient judgment scale.

By comparing the experimental items in declarative and interrogative conditions relative to their location on the gradient acceptability continuum established by the reference fillers, more reliable conclusions about the relative acceptability of BPPA constructions can be made. Since the fillers leave enough room in the upper half of the scale (A–C) for highly acceptable to slightly marked levels of acceptability, even subtle differences in declarative BPPA constructions that are obscured in binary intuitive judgments can be detected. With respect to interrogative BPPA constructions, it is of interest whether they decline all the way to the bottom of the scale in suboptimal conditions and how large the difference is to conditions that the literature considers to be grammatical.

The use of the standardized fillers in an experimental setting also has two more mundane, methodological benefits: (i) a plausibility check for the target items, and (ii) a plausibility check for participant responses.

In a typical experiment, it is advisable to construct target items that avoid the extreme points of a closed scale to prevent ceiling and floor effects. It is also advisable to exclude target items that have no unique structural representation (*word salad*) because the researcher cannot determine which structural parse is being judged (Gerbrich et al. 2019, pp. 310–11).

The E-standards are marked in several clearly determinable ways but still have a unique structural representation, whereas the A-standards contain no structural or semantic faults. This means that researchers should become skeptical if their target items fall outside the range of the standard fillers, i.e., if there are items averaging significantly better than the A-standards or significantly below the E-standards. There can be, of course, good reasons for such situations, but the results should be scrutinized closely. Target items that fall somewhere between the ranges of the standard fillers can be more clearly evaluated for their overall gradient acceptability.

The second point concerns the reliability of participant judgments. As these judgments are collected in an anonymous fashion and there are no negative consequences for incoherent or blatantly random judgments, data quality needs to be ensured at some point. Especially experiments that are carried out with compensation of some kind, be it monetary or for course credit, may create an environment where participants are not really engaged with the task and do the experiment half-heartedly. Large crowdsourcing platforms, such as Amazon's Mechanical Turk or others, have on the one hand been shown to provide usable data (Gibson et al. 2011; Sprouse 2011), but on the other hand it can always happen that participants try to complete as many tasks as possible for maximum compensation. An ethical amount of payment is a first step to avoid this, but does not guarantee accurate data. To solve this issue, the standard fillers can be evaluated for individual participants to see whether they reproduce the expected decline in mean acceptability from the A-standards to the E-standards. If the E-level averages higher than the C-level, for instance, this is a good indication that the experiment was not carried out diligently, providing a principled reason to exclude this participant from the statistical evaluation. Although the judgments for the standard fillers may not always exactly follow the expected pattern, as shown in Featherston (2020, p. 170), it is still possible to distinguish completely random judgments from those that are slightly off.

These two methodological points have shown that the standardized fillers have a valid use in experimental judgment studies in addition to the better comparability with stable levels of acceptability. They provide a more fine-grained scale of well-formedness compared to binary judgments, and also allow for a more principled conversion to traditional judgment marks, such as the question mark or the asterisk.

3.4. Interim Conclusion

In this section, I have discussed three issues that should be considered in the analysis of island-internal variation, exemplified with an evaluation of theoretical approaches to BPPA constructions. First, the relation between grammaticality and acceptability and how this relation can become problematic for theoretical conclusions about locality operations, such as *wh*-extraction. I have argued that there is nothing wrong in considering the two declarative sentences in (14) grammatical; it is, however, problematic to ignore potentially interesting differences in acceptability. Second, I have described the use of a factorial design to better describe island-internal variation in relation to the variation that is independent of extraction from the island. This design avoids potential confounds that arise if too much emphasis is placed on variation in interrogatives. Third, I have discussed how acceptability judgment tasks can benefit from the use of the standardized fillers in Gerbrich et al. (2019), from both conceptual and methodological perspectives. In combination with a factorial design that includes declarative base-structures, this allows for a principled analysis of the effects operating in specific types of islands.

In the following section, I consider existing experimental work on the acceptability of declarative and interrogative BPPA constructions and how these results compare to the issue of gradient acceptability and ramifications for the construction of licensing mechanisms for extraction.

4. Previous Experimental Investigations

The idea that not all declarative BPPA constructions are equally acceptable because the adjunct predicate is not semantically licensed in all configurations was first proposed in Brown (2015, 2016, 2017). She argues that only low-merged VP adjuncts are in the right structural configuration to allow extraction, whereas high-merged *v*P adjuncts resist extraction.[8] By hypothesis, not all types of participle adjunct predicates qualify as low-merging adjuncts to all types of matrix predicate. This means that some participle adjuncts fail to be licensed in the configuration that would allow extraction, which leads to reduced acceptability that is independent of whether extraction takes place or not. Brown (2017) formulates this as a distinction between the semantic licensing conditions on the low-merging adjunct and the syntactic licensing conditions for extraction. For the semantic licensing conditions of low-merging adjuncts, she suggests that the temporal interval of the matrix predicate should be properly included in that of the adjunct predicate, which works best if the adjunct predicate is atelic and the matrix predicate telic; this is essentially the generalization formulated in Truswell (2007). Kehl (2021) goes in a similar direction by proposing a set of semantic compatibility and syntactic complexity criteria that determine the acceptability level of the declarative BPPA construction, taking into account the properties of the host predicate. Both approaches share the common assumption that there is a principled relation between acceptability differences in interrogatives and the corresponding declaratives.

Brown (2017) shows experimentally that there are significant effects of transitivity in declarative BPPA constructions, and that this effect does not interact with the presence vs. absence of a gap: thus, the relative acceptability difference between the intransitive (24a) and the less acceptable transitive (24b) is the same as that between the corresponding declaratives in (25a) and (25b).

(24) a. Which tune did Monica arrive whistling? > more acceptable

b. Which tune did Julia pick the candidates whistling?

(Brown 2017, p. 119)

(25) a. Lucy arrived whistling the national anthem. > more acceptable

b. Mary picked the candidates whistling the national anthem.

(Brown 2017, p. 119)

What this means is that transitivity shows an effect on acceptability but does not determine how strongly extraction affects acceptability. This result is unexpected in the framework proposed by Borgonovo and Neeleman (2000).

In addition, Brown (2017) shows that telicity only has a significant effect for intransitive matrix predicates, i.e., for unergatives and unaccusatives. Transitive atelic activities and transitive telic accomplishments do not show a similar sensitivity. This is not predicted by the event-semantic account in Truswell (2007). For example, the telic transitive sentence in (26a) is equally acceptable as the atelic transitive in (26b), but the telic intransitive in (27a) is more acceptable than the atelic intransitive in (27b); the same obtains for the corresponding interrogatives. Similar observations are found in Kehl (2021).

(26) a. Mary picked the candidates whistling the national anthem.

b. Sophie finished sketches whistling the national anthem.

(Brown 2017, p. 119)

(27) a. Lucy arrived whistling the national anthem. > more acceptable

b. Lucy shivered whistling the national anthem.

(Brown 2017, p. 119)

Brown (2017) concludes that transitivity is a key factor in determining the acceptability of declarative and interrogative BPPA constructions;[9] she also concludes that the relation between acceptability contrasts in declaratives and interrogatives should be taken seriously. These results fit her two-component model with independent licensing conditions for the adjunct and extraction operations. The complex acceptability pattern observed for interrogative BPPA constructions in the literature can be traced back to similar differences in declaratives, obviating the need for additional licensing mechanisms that are tied to extraction.

Similarly, Kehl (2021) reports that telic matrix predicates have an advantage over atelic ones (experiments 1 and 2) and that unaccusative matrix predicates are judged as more acceptable compared to unergatives and transitives (experiment 4); in none of the experiments, however, do these factors interact with extraction, so that the acceptability differences in interrogatives can be reliably predicted from identical contrasts in declaratives. These results obviate the requirement for additional syntactic or semantic licensing conditions for extraction as postulated in Borgonovo and Neeleman (2000) and Truswell (2007). For example, there are already significant differences between declarative conditions with telic and atelic matrix predicates, respectively, seen in (28). To be precise, the relative difference is exactly the same as in the interrogatives in (29), as the telicity of the matrix predicate does not interact with the presence or absence of extraction.

(28) a. John arrived whistling a funny song. [telic matrix predicate]
b. John worked whistling a funny song. [atelic matrix predicate]

(29) a. What did John arrive whistling? [telic matrix predicate]
b. What did John work whistling? [atelic matrix predicate]

Additionally, the same contrasts are obtained in relativizations such as (30), which are closer in form to the attested examples in Santorini (2019). A comparison of declarative, interrogative, and relativization BPPA constructions shows that the effect of telicity remains the same across these sentence types, but the overall acceptability is shifted: declarative BPPA constructions are generally more acceptable than relativizations, which, in turn, are more acceptable than interrogative BPPA constructions. This points towards the fact that different types of long-distance dependencies require different degrees of processing effort.

(30) a. This is the song that John arrived whistling. [telic matrix predicate]
b. This is the song that John worked whistling. [atelic matrix predicate]

Similar results are obtained for the distinction between unaccusative, unergative, and transitive matrix predicates; this points towards the fact that the proposals in Borgonovo and Neeleman (2000) and Truswell (2007) are not related to extraction from the adjunct. From an architectural perspective, it is easier to include a condition on the possibility of L-marking along the lines of Borgonovo and Neeleman (2000) instead of making core syntactic operations sensitive to semantic factors; whether an event-semantic approach to acceptability differences in declaratives fares better than one based on the grammatical verb type of the matrix predicate remains to be seen, but both are most likely related to how complex the resulting BPPA construction is for the parser to interpret and how plausible the complex event described there is; see also Chaves and Putnam (2020) for similar points. It is probably the case that Truswell (2007) is on the right track concerning the influence of event structure, even if this factor does not seem to depend on the presence or absence of extraction.

Several experiments in Kehl (2021) also show that there are considerable differences between declarative conditions, which are not directly predicted in Borgonovo and Neeleman (2000) or Truswell (2007), again pointing to the importance of considering the relative acceptability of the underlying declaratives instead of only their grammaticality. These

differences in declaratives can be captured in the comparison with the standardized reference fillers from Gerbrich et al. (2019): in most of the reported experiments, there is a contrast between the more acceptable declarative conditions, which are located between the A- and the B-level of the reference fillers, and the less acceptable declarative conditions with judgments clearly below the B-level and sometimes closer to the C-level. This shows that these differences are not too subtle to be irrelevant, or "unremarkable" as Truswell (2007, p. 1373) puts it.

The declarative counterparts of BPPA constructions are also compared to interrogatives in Kohrt et al. (2018), who do not find evidence for the theoretical claims about the factor agentivity in Truswell (2011), but crucially also no interaction of their factor ±extractable with extraction (their experiment 1). Against the predictions from Truswell (2007) and Truswell (2011), they do not find significant effects of verb type distinctions between extractable *arrive*-type verbs and non-extractable *work*-type verbs; see the example items in (31a). The only significant effect they find is between declaratives (31a) and interrogatives (31b), which is the predicted negative effect of extraction on acceptability.

(31) a. John wondered whether his best friend {worked/arrived} at the office drinking some coffee late this afternoon.

b. John wondered which coffee his best friend {worked/arrived} at the office drinking __ late this afternoon.

(Kohrt et al. 2018)

The lack of a significant effect of whether the matrix predicate is a suitable predicate for extraction may partially be caused by their assignment of event types to either extractable or non-extractable conditions: they include states in the extractable category and accomplishments in the non-extractable category, which is in line with the claims about agentivity in Truswell (2011), but is problematic from the observations about telicity in Truswell (2007) and the possibility for accomplishments to allow extraction when the adjunct specifies the causal component of the accomplishment, which is explicitly acknowledged in Truswell (2011).

The experimental evidence provided by Brown (2017) and Kehl (2021) supports the hypothesis that the factors identified in the literature do not influence the strength of extraction from the adjunct; there is no need to postulate additional licensing mechanisms to evade the CED. Both find that there are systematic acceptability differences in declaratives that are carried over to the interrogative structures without additional effects requiring an explanation.

5. A Model for the Acceptability of Participle Adjuncts

Once the focus of interest is shifted to a principled comparison between declarative base positions and *wh*-interrogatives, as well as the underlying acceptability differences in declaratives, the question is what causes these acceptability differences found in Brown (2017) and Kehl (2021). In this section, I will first discuss factors which influence the acceptability of (declarative) participle adjuncts; some, but not all of these factors have been discussed in the previous literature. At the end of this section, I will combine the factors into a partially weighted model for predicting the acceptability of declarative and interrogative participle adjunct constructions. This model will be conceptually based on graded and multifactorial models of acceptability such as the Decathlon Model (Featherston 2008, 2019) and the Cumulative Effect Hypothesis discussed in Haegeman et al. (2014) and Greco et al. (2017).[10] In these types of model, the violation of individual constraints show negative effects on acceptability; these constraint violations are cumulative, so that the violation of each additional constraint further decreases acceptability. I will argue that extraction from the adjunct is simply one additional negative effect that is added to the combined effects of the factors which influence acceptability in declarative BPPA constructions; crucially, the size of the extraction effect does not depend on whether other effects apply in the

declarative or not.[11] This is precisely the fundamental assumption made in Brown (2017) and Kehl (2021), which differentiates these accounts from previous approaches to extraction from adjuncts.

5.1. Transitivity: Multiple Referents Incur Independent Processing Costs

One of the factors that determines whether a BPPA construction is highly acceptable in declaratives is transitivity, i.e., whether the matrix predicate selects one or more arguments. Brown (2017) finds that transitivity is a relevant factor because it determines whether telicity has an effect at all, shown by an interaction of the two factors in her experiments. For transitive predicates, it is not important whether it is an atelic activity or a telic accomplishment, but intransitives are sensitive to the unergative–unaccusative distinction, with unaccusative achievements being more acceptable than unergative activities. This result is also found in Kehl (2021, experiment 4), where unaccusatives have a general advantage over unergatives and transitives, which are not differentiated between telic and atelic.

An additional observation made in Kehl (2021), based on the discussion in Borgonovo and Neeleman (2000), is that the nature of the second argument is important: reflexive objects as in (32a) and subjects of resultative constructions as in (32b) behave differently than prototypical transitive predicates with two distinct discourse referents, as in (32c).

(32) a. John hurt himself [trying to fix the roof]. [reflexive]
b. John drove Mary crazy [talking about his sacking]. [resultative]
c. John cut Bill [carving the turkey]. [transitive]
(Borgonovo and Neeleman 2000, p. 211)

Here I will not go into a detailed discussion why resultative constructions differ from transitives; see Winkler (1997), Rothstein (2017), and Hu (2018) for discussion of how the subject of the resultative is assigned its θ-role. Incidentally, Borgonovo and Neeleman (2000, p. 212) observe that extraction from the adjunct in (32b) is ungrammatical, whereas Truswell (2007, 2011) considers this a prime example of transparent accomplishments; this emphasizes the need to investigate this type of matrix predicate in more detail.

In more general terms, a second argument increases complexity in the BPPA construction, also because potential control conflicts of the adjunct predicate need to be resolved: in a transitive sentence, the adjunct can be controlled by both the subject and the object of the matrix clause, which increases the amount of processing to resolve this ambiguity. Some event types show restrictions in their control possibilities (Rapoport 2019; Simpson 2005), but then the parsing of the wrong control orientation should lead to even lower acceptability.[12]

The observation that transitivity in general incurs drops in acceptability independently of extraction operations is also made in Jurka (2010, 2013), Polinsky et al. (2013), and Konietzko (2021); they all find that predicates which select a second argument are slightly less acceptable than intransitives (unergatives and unaccusatives) in declarative structures. Polinsky et al. (2013, p. 296) refer to this as a 'transitivity penalty', which is probably caused by the processing effort to parse the second argument. Similar effects of transitivity are also discussed in relation to extraction in Dependency Locality Theory (Gibson 1998, 2000), which also offers an explanation for the behavior of transitives; I follow Polinsky et al. (2013) in assuming that the effects of transitivity are not exclusive to sentences with extraction.

The negative effects of transitivity make the prediction that the more arguments are selected by the matrix predicate, the higher the processing effort required of the parser, with at least some effect on acceptability. Thus, I predict a relative decline in the acceptability of the sentences in (33), even if all structures might receive a grammatical binary judgment:

(33) a. John arrived singing an obscene song. [intransitive]
b. John offended Mary singing an obscene song. [transitive]
c. John gave Mary a letter singing an obscene song. [ditransitive]

The full paradigm of transitivity thus ranges from purely intransitive to reflexive transitive, resultative, transitive, and, finally, ditransitive. It is also possible that not only the number of arguments, but also other factors play a role; this could be formulated in terms of the multi-faceted definition of the transitivity continuum in Hopper and Thompson (1980). An additional problem that arises in ditransitives is that there is a potential orientation ambiguity for the participle adjunct depending on its lexical content: the orientation can be shifted towards the direct object, as in (34), and is sometimes the preferred interpretation.

(34) John$_i$ gave Mary$_j$ a letter$_k$ [lying on the table]$_k$.

In the interrogatives corresponding to (33), the contrast between the intransitive and the (di-)transitive structures is noticeable, but the ditransitive is even worse than the transitive. This is not directly reflected in the binary judgments in (35), but should be visible in a judgment study. The low acceptability of the ditransitive structure (35c) carries over to the alternative ordering in the double object construction in (35d).

(35) a. What did John arrive singing __ ? [intransitive]
b. *What did John offend Mary singing __ ? [transitive]
c. *What did John give Mary a letter singing __ ? [ditransitive I]
d. *What did John give a letter to Mary whistling __ ? [ditransitive II]

Chaves and Putnam (2020, p. 15) point to the fact that optional transitivity may confound the intended interpretation of interrogative BPPA constructions because the *wh*-phrase may be linked to a gap in complement position of an optionally transitive matrix predicate instead of the complement position of the adjunct; see also Staub (2007) and Ness and Meltzer-Asscher (2019).[13] This ambiguity is shown in (36), where potential optional gap sites are indicated by underscores in parentheses.

(36) What$_{i/j}$ did John walk (__$_i$) whistling (__$_j$)?
a. John walked the dog whistling.
b. John walked whistling a funny song.

An ambiguous parse with gap position after the matrix predicate can be avoided by restricting adjunct predicates to obligatorily transitive predicates, such as *proclaiming*, as in (37). Here the gap site after the main verb would trigger ungrammaticality because the gap after the adjunct is obligatory, here indicated by the lack of parentheses around the gap site following the adjunct predicate. This means that the *wh*-pronoun cannot associate with the optional potential gap site in the matrix clause. A parasitic gap reading is also possible here if the filler can be the object of both predicates; I do not discuss this possibility further here.

(37) What$_{i/j}$ did John walk (__$_i$) proclaiming __$_j$?
a. *John walked the dog proclaiming.
b. John walked proclaiming his love for Pam.

Yet another way to reduce gap site ambiguity is if a motion verb like *walk* is augmented with a directional phrase, as in (38); it is still possible that John walks his dog to the park, but this parse becomes less likely than in (37).

(38) What did John walk to the park whistling __ ?

To sum up, transitivity, even if it is optional, increases the overall complexity of the BPPA construction and thus gradually builds up hurdles for extraction. Unambiguously intransitive predicates are predicted to have an advantage over potentially transitive and

unambiguously transitive predicates; reflexive and resultative predicates occupy the middle ground because on the one hand they include a second argument, but this argument is either not directly selected by the main verb (resultatives) or is co-referential with the main verb's subject (reflexives).

5.2. Event Structure: Durativity Instead of Telicity

Another factor which has an effect on the acceptability of declarative and interrogative BPPA constructions is based on the observation that not all types of matrix predicate can be felicitously modified by an adjunct predicate. The restrictions on BPPA constructions resemble those that operate in depictive secondary predication, where likewise not all types of main verb accept depictives to the same degree (Rapoport 2019; Simpson 2005). There is an ongoing discussion whether complex adjuncts, such as BPPAs, can be analyzed as depictives, but I will assume this for the present discussion; see also Rothstein (2017, p. 3874). For example, permanent statives, as in (39a), are odd with a BPPA, whereas temporary statives, as in (39b), are more acceptable.

(39) a. ?John was blond [wearing his new sunglasses]. [permanent state]
b. John lay in bed [wearing his new sunglasses]. [temporary state]

The difference between these types of states is that temporary states have an event variable, which permanent states lack (Rapoport 1993, p. 173). Permanent states are property ascriptions whereas temporary states are predicated of the subject for a temporal interval that allows delimitation. This distinction also shows up in the corresponding interrogatives in (40):

(40) a. *What was John blond [wearing __]?
b. What did John lie in bed [wearing __]?

Since both permanent and temporary states are atelic, these acceptability differences are problematic for the telicity-based account in Truswell (2007) and Brown (2017), as well as for the reflexivity account in Borgonovo and Neeleman (2000).

A telicity requirement is also problematic for purely punctual achievements like *appear*, which should be ideal candidates for a temporal inclusion relation in Brown (2017); still, these predicates are degraded in interrogatives, as seen in (41):

(41) a. John appeared wearing a beautiful bespoke suit.
b. *What did John appear [wearing t]?

(Truswell 2007, p. 1374)

Similar observations can be made for verbs such as *notice* and other perception verbs. The question is whether this carries over to the declarative counterparts; as far as I am aware, this has not been directly tested in a controlled experiment. What permanent states and purely punctual achievements have in common is that both fail to felicitously appear in the progressive, as seen in (42a) and (42b). Crucially, temporary states are fine with the progressive, shown in (42c).

(42) a. ?John is being blond.
b. ?John is appearing.
c. John is lying in bed.

In terms of Rothstein (2004), punctual achievements and many perception verbs such as *notice* fail to appear in the progressive because the progressive cannot target an interval preceding the culmination point. The situation is different in cases similar to *arrive*, where the preceding interval can be conceptualized as the path component that leads up to the

arrival. With *appear*, the perspective is different: it is inherently external to the appearing entity, whereas *arrive* allows a conceptualization from the perspective of the arriving entity. This is a first indication that telicity alone makes the wrong predictions in these cases; rather, it seems that there is a certain correlation between the reported interrogative patterns and the ability to appear in the progressive.

Thus, the generalization about telicity in Truswell (2007) needs to be revised to exclude purely punctual achievements and to allow for temporary states. Instead of telicity, I argue that a first step towards a descriptive pattern is to consider the encoding of a durative subevent as relevant for acceptability, which is not the case for permanent states and punctual achievements.

5.3. Incrementality: Themes, Paths, and Properties

An exclusive focus on durativity leads to problems with the experimental results for activity main verbs in Brown (2017) and Kehl (2021): BPPA constructions with activity main verbs are less acceptable than achievements. To further constrain declarative BPPA constructions, a comparison with depictive secondary predicates shows that not all activity main verbs license a depictive, as shown in (43). The pattern is more difficult to capture than that of permanent and temporary states or punctual achievements, but if the BPPA construction can be analyzed as depictive secondary predication, similar effects can be expected there as well. It is also noteworthy that the addition of an object in (43c) ameliorates the modification of *draw* by a depictive.

(43) a. Jones danced/lectured drunk.
b. *Jane laughed/drew drunk.
c. Jane drew pictures drunk.

(Rapoport 2019, pp. 434–35)

The distinction between *draw* and *draw pictures* in (43b) and (43c) also shows up in BPPA constructions, where the bare form in (44) is degraded in the interrogative; as noted above, the experimental evidence in Brown (2017) and Kehl (2021) suggest that the declarative counterparts are also less acceptable than sentences with achievement main verbs.

(44) a. I work listening to music.
b. *What do you work [listening to t]?

(Truswell 2007, p. 1373)

The sentences improve in the presence of a direct object, seen in (45). This is contrary to the expectations derived from transitivity in the previous subsection, but suggests that some form of temporal delimitation may be a factor contributing to acceptability, without leading to a telicity requirement.

(45) a. Mary worked on her thesis drinking coffee.
b. What did Mary work on her thesis drinking __ ?

All the acceptable depictive constructions in (43) involve an activity predicate that is in some sense delimited, but still atelic. A specific dance or a lecture have a specified duration, and the drawing of pictures can be measured by the amount of pictures produced, whereas laughing and drawing in the sense of aimlessly doodling are not delimited in the same sense. It could be argued that this type of delimitation is connected to the concept of incremental themes (Dowty 1979): a lecture, pictures, and working on a thesis can be measured against a scale of progress, similar to the incrementality of eating one, two, or three apples. The analogy to incremental themes also extends to the domain of motion, which also come in incremental and non-incremental forms. As noted in Dowty (1979),

Tenny (1995), and Borgonovo and Neeleman (2000), unergative manner of motion verbs like *walk* behave differently when they are followed by a directional PP like *to the station*; this PP introduces a path component that can be measured similar to incremental themes. The effect is shown in (46):

(46) a. ?Mary walked whistling a funny song.
b. Mary walked to the station whistling a funny song.

Incrementality also extends to properties, which captures cases such as (47), where the degree of being scared increases with the progress through the movie (the gradual reading of this sentence probably comes from the durative character of the adjunct predicate, but this discussion is outside the scope of this paper).

(47) John got scared watching a horror movie.

Similar effects of incrementality are seen with semelfactive main verbs such as *jump* in (48), where a particle inducing iterativity and thus durativity has a positive effect in interrogatives. A possible factor in addition to transitivity and durativity could thus be the potential of the event described by the main verb to be measurable or quantifiable in some sense.

(48) What did she jump *(around) [singing t]? (Truswell 2007, p. 1361)

Taken together, there is at least some evidence that purely temporal inclusion of the matrix interval within the interval of the adjunct predicate is not able to account for the full data pattern, which casts doubt on the scale amalgamation process suggested in Brown (2017). The overall picture emerging from this discussion is that it is unlikely that there is a single factor which determines whether a given main verb will be highly acceptable with a BPPA. This bears close similarity to the multiple factors which influence performance and acceptability along the lines of Chomsky (1965), suggesting that the acceptability of declarative BPPA constructions is a matter of syntactic and semantic complexity and compatibility criteria instead of strict syntactic licensing requirements.

5.4. Combining the Factors into an Acceptability Model

Based on the theoretical discussion of the relation between acceptability in declarative and interrogative BPPA constructions in Brown (2017) and the evidence supporting it, Kehl (2021) develops a model that captures this relation; this model includes factors that differ from those in Brown (2017) and other approaches. The main focus is on the fact that the factors which operate in interrogatives are also visible in declaratives. Extraction simply acts as an additional factor that is independent of the individual decreases in acceptability resulting from other factors, such as transitivity or durativity. The model can be summarized as follows:

(49) Model for the acceptability of BPPA constructions:
i. Determine the acceptability of the declarative sentence; factors: transitivity, durativity, incrementality
ii. Determine the acceptability of the interrogative sentence by adding the processing costs of extraction to the result of (i)

In the first stage of the model (49i), the factors discussed above influence the acceptability of the BPPA construction: transitivity will decrease acceptability because more arguments require more processing effort. Durativity and incrementality work similarly: the absence of a durative subevent, i.e., for permanent states and purely punctual achievements, decreases acceptability, as does the absence of a delimited or incremental meaning component. Transitivity is most likely a result of increased processing effort, but durativity

and incrementality are semantic factors which seem more related to the conceptual felicity of the situation described in the sentence. Kehl (2021) collects durativity and incrementality under the term *semantic compatibility*.[14] In contrast to these factors, transitivity can be captured in syntactic terms, but the reason that transitivity matters is more likely to be found in relation to ease of processing and the ambiguity between transitive and intransitive uses of the verb in question.

The second stage of the model (49ii) adds the cognitive cost of establishing a dependency (Wagers 2013); this cost is most likely higher than into other domains, such as subcategorized complements, in line with the CED.[15] As this dependency formation is more demanding than a declarative sentence, this results in decreased acceptability. Crucially, the application of extraction and the resulting decreases in acceptability are independent of the factors which determine acceptability in the declarative: in a sense, extraction is blind to these factors. This is compatible with the independence of syntactic operations from purely semantic properties of the sentence (Brown 2017).

With respect to the relative weight of the factors that affect acceptability in declarative BPPA constructions, the previous experimental work on this construction in Brown (2017), Kohrt et al. (2018), and Kehl (2021) does not directly allow conclusions. The negative effect of transitivity is observed and isolated as a key factor in Brown (2017) and is in agreement with the transitivity penalty discussed in Polinsky et al. (2013). Scalar change and durativity are more complex to evaluate because the previous experimental work has focused on the telic–atelic distinction to check the predictions of Truswell (2007), but this distinction does not directly map to the factors discussed here. The complex interactions of these factors should be addressed in future experimental research. Based on the experimental results from Brown (2017) and Kehl (2021), it is possible to assign a preliminary weighting to this model: the effect of extraction is much stronger than that of durativity, incrementality, or transitivity. This observation connects to the discussion above about the subtle acceptability differences in declarative BPPA constructions, which run the risk of being considered irrelevant, especially if the focus of the approach in question is in grammaticality rather than acceptability. The acceptability model can be graphically represented as in Figure 3, taken from Kehl (2021).

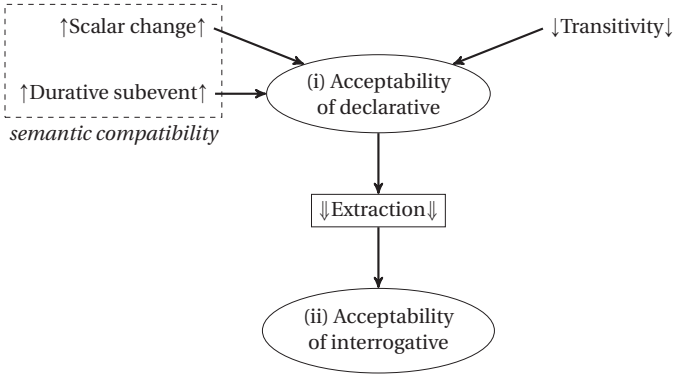


Figure 3. Acceptability model for BPPA constructions from Kehl (2021, p. 284); the factor *scalar change* corresponds to incrementality in this paper. Upwards arrows indicate a positive effect on acceptability, (double) downward arrows a negative effect.

This illustration shows the positive effects of durativity and incrementality with upward arrows, as well as the negative effect of transitivity with downward arrows; double downward arrows on the factor extraction indicate that this effect is stronger than the others. The central characteristic of this model is that it incorporates the relation between declarative and interrogative acceptability as formulated in Brown (2017), which is stated in Kehl (2021) as the independence of extraction from the factors operating in declaratives.

This model accounts for the sometimes subtle acceptability differences in declarative BPPA constructions, as well as the central factors isolated for participle adjunct islands in Borgonovo and Neeleman (2000) and Truswell (2007). At the same time, however, this model is conceptually simpler because the extraction operation remains blind to semantic characteristics of the sentence in question.

The model captures the following judgment differences discussed in the literature: (i) the advantage of telic over atelic matrix predicates due to scalarity (50i), (ii) the oddity of punctual matrix predicates because the latter do not satisfy durativity (50ii), (iii) the improvement with path scales and incremental themes for atelic matrix predicates because they introduce a scalar meaning component (50iii), and (iv) the effect of the number of arguments selected by the matrix predicate as a reflex of transitivity (50iv). If these contrasts can be shown to be observable in declaratives as well as interrogatives alike, this supports the predictions of the factorial acceptability model.

(50) i. What did John arrive/*work whistling __ ? [scalarity]
ii. *What did John appear/notice whistling __ ? [durativity]
iii. What did John work *(on his thesis) whistling __ ? [scalarity]
iv. What did John hurt himself/*Bill trying to fix __ ? [transitivity]

Not all of these contrasts have been tested experimentally in the literature: the contrast in (50i) is the one that most of the existing literature focuses on, e.g., Brown (2017), Kohrt et al. (2018), and Kehl (2021). Likewise, transitivity effects as in (50iv) are to a certain extent explored in these studies, but further studies are required to see where reflexive and resultative matrix predicates lie in relation to intransitive and transitive sentences. The contrasts between purely punctual and extendable achievements in (50ii) as noted in Truswell (2007) and the precise effect of an added scalar meaning in cases like (50iii) also require additional work.

This acceptability model focuses on simple declarative and interrogative BPPA constructions, but it can also be modified to include other sentence forms, such as relativization or topicalization; these sentence forms also encode unbounded dependencies, but are not interrogative (Chaves and Putnam 2020). It can thus be expected that they do not show the same degree of decreased acceptability as the *wh*-interrogatives focused on in this article, which is also indicated in the data reported in Abeillé et al. (2020) and Liu et al. (2022). Compare the declarative BPPA construction in (51) with the different types of dependencies in (51a)–(51c).

(51) John arrived [whistling an annoying song]. [declarative]
a. Which song did John arrived [whistling $\__i$]? [*wh*-interrogative]
b. This annoying song$_i$, John arrived [whistling $\__i$]. [topicalization]
c. I hated the song$_i$ (that/which) John arrived [whistling $\__i$]. [relativization]

Initial evidence that relativization leads to a generally smaller decrease in acceptability than bare *wh*-interrogatives is given in (Kehl (2021) [experiment 1]). This might be related to a better match between the information-structural status of the adjunct constituent from which extraction takes place and the discourse function of relativization, as proposed in Abeillé et al. (2020). The visualization of the acceptability model in Figure 3 can be generalized by adding more extraction types than just *wh*-extraction, and by linking these different types of dependency formation to separate acceptability levels; this is shown in Figure 4, where relativization and topicalization are allowed for negative effects on acceptability that are not necessarily identical to that of *wh*-extraction. I will have to leave the relative magnitude of these effects for future experimental research. The underlying hypothesis remains that the contrast between matrix verbs such as *arrive* and *work* can be observed equally across these different dependency types; this assumption follows the

argumentation in Chaves and Putnam (2020) that the pragmatic felicity of the underlying proposition has a strong role to play in island effects and extraction asymmetries.

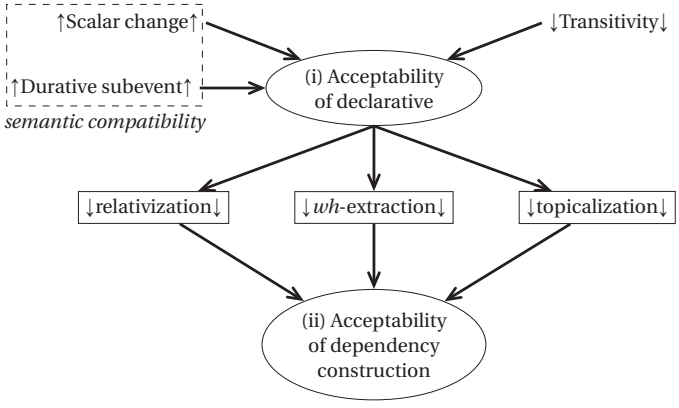


Figure 4. Adapted acceptability model for BPPA constructions in different sentence forms; the acceptability of a dependency construction is modeled as the acceptability of the underlying simple declarative plus the effect of establishing a dependency construction. Upwards arrows indicate a positive effect on acceptability, downward arrows a negative effect.

Another important issue is how strong the factors of the acceptability model are affected by variation in speaker judgments. So far, I am not aware of experimental studies that explicitly take this factor into account. There are studies on the related phenomenon subject islands investigating whether judgments improve depending on presentation order: Chaves and Dery (2019) report that judgments improved if the item was presented later in the experiment, suggesting that there is a satiation effect and that the initially low acceptability judgment improves with repeated exposure as a type of learning effect. If violations of the subject condition can improve over time, it seems plausible that the type of semantic mismatches resulting from scalarity and durativity can also improve with repeated exposure, but this requires further investigation.

In conclusion to the factors related to acceptability in the BPPA construction and the model proposed in Kehl (2021), it seems that Truswell (2007) is not right in his claim that declarative BPPA constructions which do not meet his extraction condition are unremarkable. The exact opposite holds: acceptability differences in declaratives resulting from a variety of different factors are the key determinants of acceptability in interrogative BPPA constructions, and it is not the extraction operation that triggers these differences in interrogatives.

6. Converging Evidence for the Relevance of Acceptability Differences in Declaratives

More recent work agrees about the relevance of potential acceptability differences in declaratives for the acceptability of movement constructions. The proposals diverge slightly in the source of such differences, but the focus has shifted from purely syntactic explanations towards more interface-based ones. Transitivity as a processing-related complexity criterion and event structure as a semantic notion have been the focus of this article.

Similar conclusions about extraction-independent effects of processing-related complexity on acceptability are drawn for the apparent licensing of island-violating extractions in so-called parasitic gap (PG) environments in Culicover and Winkler (2022); they trace back the ameliorating effect ascribed to parasitic gaps to complexity differences between (declarative) PG and non-PG constructions. The former are more acceptable because they are less complex for processing due to the fact that one less referentially distinct argument is encoded. In the contrast in (52), the additional gap in the matrix clause in (52b) leads to the fact that there is only one discourse referent in the sentence, whereas there are two

in (52a). There is, thus, an underlying difference in complexity that pushes (52a) below a threshold for grammaticality, which is not the case in (52b). The parasitic gap is indicated by *pg* in this example.

(52) a. *a person who$_i$ [talking to t_i] about this would prove to the Mayor that there is a problem

b. a person who$_i$ [talking to pg_i] about this would prove to t_i that there is a problem
(Culicover and Winkler 2022, p. 2)

Although the two corresponding non-extraction sentences in (53) are certainly both grammatical, (53a) is more complex than (53b) from a processing perspective because an additional discourse referent needs to be processed. Whether this results in noticeable acceptability differences is a question that is outside the scope of this paper, but can explain the strong judgment difference reported for (52) by Culicover and Winkler.

(53) a. Talking to *person X* about this would prove to *person Y* that there is a problem.

b. Talking to *person X* about this would prove to *person X* that there is a problem.

The conclusions in Culicover and Winkler (2022) are very similar to that discussed in this article: there is no requirement for a dedicated licensing or repair mechanism associated with parasitic gaps; a sufficient description of the underlying complexity differences is sufficient to explain why PG constructions are more acceptable than the non-PG construction. Culicover and Winkler (2022) also discuss the important distinction between grammaticality and acceptability that can be used to provide a comprehensive explanation of the patterns detected for parasitic gaps in the literature.

Another set of factors comes from the interface of syntax with pragmatics: Chaves and Putnam (2020) point out that apparent grammaticality contrasts in syntactically marked constructions, such as *wh*-questions, often have their origin in sometimes subtle pragmatic differences that are unrelated to the formation of the marked construction. They propose a largely pragmatic approach to most island domains by arguing that the low acceptability can often be traced back to issues of relevance and salience: if the island domain is not salient or relevant, acceptability contrasts in unmarked constructions can arise and evoke the impression of stronger grammaticality contrasts in marked constructions. This is captured in the Relevance Presupposition Condition (RPC):

(54) RELEVANCE PRESUPPOSITION CONDITION: the referent that is singled out for extraction in a UDC must be highly relevant (e.g., part of the evoked conventionalized world knowledge) relative to the main action that the sentence describes. Otherwise, extraction makes no sense from a Gricean perspective, as there is no reason for the speaker to draw attention to a referent that is irrelevant for the main contribution of the sentence to the discourse. (Chaves and Putnam 2020, p. 206)

The contrast in (55) is given as an example of this, but the grammaticality difference is unrelated to extraction:

(55) a. What did you read a book about?

b. *What did you drop a book about?
(Chaves and Putnam 2020, p. 207)

It has been noted as early as Kuno (1987) that the corresponding declaratives already show a noticeable acceptability difference; this is shown in (56):

(56) a. Speaking of Napoleon, I just read a book about him.

b. ?Speaking of Napoleon, I just dropped a book about him.

(Chaves and Putnam 2020, p. 205)

The reasoning to explain these independent acceptability differences is along the following lines: verbs evoke certain conceptualizations when they are encountered by the parser, and some meaning components are more easily accessible than others. Reading a book evokes the concept of a topic covered by the book, which is relevant information. However, the topic is not as relevant and easily evoked when a book is dropped (Chaves and Putnam 2020, p. 207). This has clear ramifications for acceptability in marked constructions, but may not be as clear in unmarked ones.

For BPPA constructions, the predictions of the RPC predict that adjuncts which supply relevant information invoked by the event described in the matrix predicate can be targeted by extraction. This serves as an explanation for the relative acceptability of cases where the adjunct describes the cause of the matrix predicate, as in (57a). The distinction between non-causal adjuncts discussed in Truswell (2007, 2011), as in (57b) and (57c) is less clear, but it could be argued that telic predicates like *arrive* are informationally light, so that the adjunct can be analyzed as relevant in the sense of Chaves and Putnam (2020); atelic predicates such as *work*, on the other hand, can be argued compete with the adjunct in terms of which information is more relevant, so that the extraction is not licensed by the RPC.

(57) a. What did Peter drive Mary crazy whistling __ ?

b. What did Peter arrive whistling __ ?

c. *What did Peter work whistling __ ?

The acceptability model discussed in the previous section is not mutually exclusive with the RPC; the generalizations in the model could be seen as factors that influence the relevance of the adjunct compared to the matrix predicate and hence have an effect on the acceptability of extraction. I agree with Chaves and Putnam (2020, p. 230) that "extraction from such island environments is contingent on the proposition itself, rather than strictly on its syntax". This captures the idea in the model that the factors described by the generalizations show effects that are independent of extraction.

There exists a number of experimental studies that test the relation between declaratives and interrogatives in related phenomena:[16] for example, Chaves and King (2019) find a strong correlation between plausibility ratings for declaratives and acceptability of subextraction from objects, indicating that plausibility ratings act as a predictor of acceptability that is not modulated by extraction. However, Chaves and Putnam (2020) report on another experiment investigating extraction from tensed adverbial clauses, where they do not find a correlation between declarative and interrogative acceptability, meaning that the latter is not reliably predicted by the former. In such cases, it is reasonable to assume that there is another factor which distorts the relation, similar to the factorial definition of island effects in Sprouse and Hornstein (2013). The effects of tensed adjuncts are also discussed from a theoretical perspective in Truswell (2011, pp. 175–79) and experimentally investigated in a cross-linguistic study in Müller (2019). Abeillé et al. (2020) examine relativization from subjects and objects, with the result that extraction from subjects is actually better than extraction from objects, contrary to the predictions of locality constraints such as the CED, which do not discriminate between different types of extractions; this points towards the conclusion that not all extractions function alike, and that the discourse functions of the extraction operation and the extracted element should be included in an analysis.

These brief glances beyond the scope of this paper show that theory development is well advised to take subtle acceptability differences in declaratives seriously in the discussion of licensing mechanisms for movement. Differences in processing complexity, semantic compatibility, and pragmatic characteristics can affect canonical word orders to

such a degree that the application of movement operations invokes the impression of strong grammaticality differences.

7. Conclusions

In this article, I have emphasized the importance of the underlying declarative sentences in the discussion of extraction from participial adjunct islands. Once the distinction between grammaticality and acceptability is taken seriously, it becomes possible to explain the acceptability differences in interrogatives by examining potential acceptability differences in the declarative counterparts. The result is an approach to extraction from adjunct islands that does not require additional and complicated licensing machinery as in the theories presented in Borgonovo and Neeleman (2000) or Truswell (2007, 2011). The approaches in Brown (2017) and Kehl (2021) both emphasize the relevance of acceptability differences in declarative BPPA constructions and propose factors to capture the acceptability variation independently of extraction. I have discussed three factors that are of interest in these accounts: the notion of transitivity, expressed in the number of arguments directly selected by the main verb, the event structure of the main verb, as well as the encoding of an incremental measure scale in the matrix predicate. The effect of transitivity can be described as a processing advantage of verbs with lower transitivity: more arguments to be processed incurs processing costs that can be reflected in acceptability. As far as event structure is concerned, I have argued that a simple telicity requirement, as postulated in Truswell (2007) and Brown (2017) is insufficient to explain the low judgments observed in the literature (e.g. Truswell 2007, p. 1370) for extraction from BPPA constructions with purely punctual matrix predicates, such as *appear* and the relatively acceptable judgments with temporary stative predicates, such as *lie in bed* (Truswell 2011, pp. 158–59). One of the key components isolated in the discussion is durativity instead of telicity, even if further factors need to be taken into account in order to explain the low acceptability with activity matrix predicates. The last factor is that of incrementality, where the progression of the matrix predicate can be measured against an incremental scale, formulated either as paths, incremental themes, or property values. Together, these factors provide a first set of tools to capture the acceptability differences in declarative and interrogative BPPA constructions without the need for additional, complex licensing mechanisms.

A final, more programmatic note about the nature of so-called 'island constraints' such as the CED: there is recent evidence that not all extraction types show the same effects in CED-violating operations, and that the magnitude of the extraction effect also depends on other factors of the island domain. For example, Abeillé et al. (2020) have shown that relativization has a different effect than *wh*-extraction in subject islands, which is hard to explain in pure syntactic terms such as the CED; similar observations are reported in Kehl (2021) for *wh*-extraction and relativization from BPPA constructions, who finds that relativization from BPPAs is more acceptable than *wh*-extraction, and that the aspectual classes of the matrix and adjunct predicates have identical effects, as in declaratives and interrogatives. Additionally, experimental work in Müller (2019) suggests that some adverbial clauses are harder to extract from than others, involving factors such as adverbial clause type and tense-marking. It would appear that the notion of categorical extraction constraints, such as the CED, should be critically evaluated: are such constraints really binary in core syntactic terms, meaning that the grammar can compute the extraction only in one but not in another configuration? Or is this the same type of overgeneralization that has been shown here to be problematic for accounts like Borgonovo and Neeleman (2000) and Truswell (2007)? This is a general problem faced by binary or categorical models of grammar because they are at risk of glossing over subtle acceptability differences in favor of broad general predictions; a graded model of grammar such as the Decathlon Model (Featherston 2008, 2019) has the flexibility of assigning individual decreases in acceptability to different operations from minimally different constructions, so that these effects can be individually quantified and summed up to predict acceptability in a wider range of configurations than the categorical predictions of the CED. The upshot from this brief

discussion is that there are good reasons to assume that extraction from some structural domains is harder than extraction from others, as captured in the original formulation of the CED; whether this is due to derivational or structural factors (competence-based) or the result of increased processing complexity (performance-based) is beyond the scope of this article. I leave the details of such an analysis of island constraints to future research and conclude here that BPPA constructions are an interesting showcase of island-internal variation that can be fruitfully employed to dive deeper into the nature of acceptability and its relation to intuitively observed grammaticality patterns in island constructions.

Funding: This article is the result of my work in the A7 project "Focus and extraction in complex constructions and islands" of the Collaborative Research Center SFB 833 at the University of Tübingen, which was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—SFB 833—Project-ID 75650358.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This article greatly benefited from discussion with Susanne Winkler, Sam Featherston, Peter W. Culicover, and Andreas Konietzko. I am also grateful to the three anonymous reviewers for their detailed comments and criticism, relevant literature, and overall suggestions for improvement. All remaining errors and shortcomings are necessarily my own.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BPPA	Bare Present Participle Adjunct
CED	Condition on Extraction Domain
DP	Determiner Phrase
PG	Parasitic Gap
PP	Preposition Phrase
RPC	Relevance Presupposition Condition

Notes

1 See Stepanov 2007 for an overview of developments in the Minimalist tradition, many of which incorporate the concepts of phases (Chomsky 2001, 2008) and cyclicity (Abels 2012). There are by now several approaches which syntactically derive adjunct-internal gaps without violating syntactic principles, such as Narita (2014), den Dikken (2018), and Brown (2017).

2 I do not address in detail the questions whether the effects of the CED in its original formulation or later adaptations are the result of a grammatical constraint (competence-based in terms of Chomsky 1965) or of a conspiracy of other factors, such as complexity or plausibility (which are rooted in performance). The influence of performance-based factors on extraction operations reflects a growing body of research in this direction; see, among others, Sag et al. (2008), Hofmeister and Sag (2010), and Chaves and Putnam (2020).

3 Both accounts are based on intuitive author judgments without direct empirical validation; their predictions have since been experimentally tested with mixed results in Brown (2017), Kohrt et al. (2018), and Kehl (2021).

4 Experimental research on BPPA constructions in Brown (2017) and Kehl (2021) does not confirm this intuition because there are significant acceptability differences between different declarative conditions, which points to the imperfect alignment between grammaticality and acceptability, which I discuss in Section 3. I discuss previous experimental results on extraction from adjuncts in more detail in Section 4. The relevance of considering even small acceptability differences in declaratives and relate them to differences in non-canonical structures is now more prevalent in the literature.

5 A reviewer wonders whether working memory capacity impacts acceptability in these cases. Experimental studies on the relation of individual working memory and acceptability judgments in island phenomena, such as Sprouse et al. (2012, 2013), have so far not shown a significant correlation between these two measures; as noted by the authors, this is partially due to the method in which memory capacity is measured. Sprouse et al. (2012, p. 116) report on no interaction between the dependency length effect and participant groups with high and low working memory scores when using recall scores as a measure of memory capacity, but a significant interaction if three-back scores are used.

[6] In other words, I assume that extraction from adjuncts, such as (16), always shows a superadditive island effect in terms of Sprouse and Hornstein (2013, p. 2); since I focus on a single adjunct construction, the question is whether semantic differences further influence how strong the island effect of extraction is.

[7] If these control conditions are part of the experimental design, they can be directly included in the statistics; Abeillé et al. (2020) implement this with sub-models that isolate the effects and interactions of the different factors.

[8] This distinction is explained in purely tree-geometric terms in Brown (2017), but it could also be argued that high *v*P adjuncts are inactive from a phase perspective, perhaps because they are introduced by late merger (Stepanov 2007) and are thus entirely opaque for syntactic operations that apply earlier in the derivational cycle.

[9] An effect of transitivity is not surprising from the perspective of Dependency Locality Theory (Gibson 1998, 2000) because the dependency crosses over another discourse referent only in the transitive cases. The effect of transitivity is significant independently of extraction (Brown 2017, pp. 124–25), which is compatible with the findings in Jurka (2010, 2013) and Polinsky et al. (2013).

[10] A crucial difference between constraint satisfaction models, such as the Cumulative Effect Hypothesis and the Decathlon Model, is that these models work with cross-constructionally active constraints that are part of a speaker's competence; in contrast, the generalizations discussed in the model discussed here are conceived of as factors that contribute to semantic compatibility and syntactic complexity, both with ensuing effects on acceptability. These generalizations need not be part of a speaker's competence grammar, but can be linked to how easily a given sentence is processed, considering that two events plus the temporal and semantic relations between them need to be processed.

[11] To see whether there is a difference between extraction from the adjunct and extraction in general, a reviewer suggests that extraction from BPPAs should be compared with extraction from gerundive complements, which are very similar in their surface structure. For example, is the effect of extraction in *What type of cigars did John stop smoking __ last week?* of the same magnitude as in *What type of cigars did John arrive smoking __ last week?* and does the presence or absence of an adjunct have an effect in the corresponding declaratives?

[12] A reviewer suggests that there could be a preference to have a simpler situation in the matrix predicate when the adjunct is complex; I agree that this could be a more general explanation for the effects of transitivity with complex adjuncts. The same reviewer also points out that sentences "tend to be subject–verb–complement or subject–verb–adjunct more than subject–verb–complement–adjunct". A discussion of expectation and usage frequency goes beyond the scope of the present article, but provides a fruitful area for future research.

[13] The ambiguity between transitive and intransitive alternates is also related to the relative frequency of the two forms: some ambiguous verbs occur primarily in their intransitive uses, others in combination with a prepositional complement, as transitives, and other constructions. This may influence whether a verb is preferably parsed as intransitive or transitive. Roland et al. (2007) analyze the frequencies of such occurrences in large corpus data; for example, the verb *walk* used in (36) has a frequency ordering of PP > intransitive > transitive, whereas the verb *leave* has the frequency ordering transitive > intransitive > PP (both verbs also occur in other configurations, this is just an expository selection). I thank an anonymous reviewer for pointing me towards this corpus data. For the role of frequency data in relation to acceptability from island domains, see for example Chaves and Richter (2020).

[14] Whether these semantic factors are actually encoded in the grammar and, thus, part of a speaker's competence is a question (raised by one of the reviewers) that goes beyond the scope of this article; it seems possible that there is a considerable degree of inter-speaker variation in the judgments of these factors, which should be explicitly tested in additional studies.

[15] This means that a locality condition such as the CED still has a place in syntactic theory, but the question whether it is a categorical constraint or rather a gradual phenomenon should be investigated more closely. If it can be shown that dependencies into non-complement domains are computationally possible, i.e., locally well-formed (as exemplified for low VP-adjuncts in Brown 2017), the CED could be reduced to a processing phenomenon instead of a grammatical principle. Research in this direction is still ongoing, as in Culicover and Winkler (2018) and Culicover et al. (2022), but an interesting venue for future research.

[16] I am grateful to one of the reviewers for pointing out these relevant studies on closely related issues.

References

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204: 104293. [CrossRef] [PubMed]

Abels, Klaus. 2012. *Phases: An Essay on Cyclicity in Syntax*. Berlin and Boston: de Gruyter. [CrossRef]

Bader, Markus, and Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46: 273–330. [CrossRef]

Bader, Markus, and Jana Häussler. 2019. How to get from graded intuitions to binary decisions. In *Experiments in Focus: Information Structure and Semantic Processing*. Edited by Sam Featherston, Robin Hörnig, Sophie von Wietersheim and Susanne Winkler. Boston and Berlin: de Gruyter, pp. 183–207. [CrossRef]

Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72: 32–68. [CrossRef]

Borer, Hagit. 2005. *Structuring Sense, Vol. 2: The Normal Course of Events*. Oxford: Oxford University Press. [CrossRef]

Borgonovo, Claudia, and Ad Neeleman. 2000. Transparent adjuncts. *Canadian Journal of Linguistics/La Revue Canadienne de Linguistique* 45: 199–224. [CrossRef]
Brown, Jessica M. M. 2015. Apparently semantically-motivated extraction in an autonomous syntax. *Cambridge Occasional Papers in Linguistics* 8: 6–23.
Brown, Jessica M. M. 2016. Blackholes and subextraction from adjuncts in English and Norwegian. In *Proceedings of the Chicago Linguistic Society (CLS)*. Chicago Linguistic Society. Chicago: Chicago Linguistic Society, vol. 51, pp. 67–81.
Brown, Jessica M. M. 2017. Heads and Adjuncts: An Experimental Study of Subextraction from Participials and Coordination in English, German and Norwegian. Ph. D. thesis, University of Cambridge, Cambridge, UK.
Brown, Jessica M. M., Gisbert Fanselow, Rebecca Hall, and Reinhold Kliegl. 2021. Middle ratings rise regardless of grammatical construction: Testing syntactic variability in a repeated exposure paradigm. *PLoS ONE* 16: 1–26. [CrossRef]
Chaves, Rui P., and Adriana King. 2019. A usage-based account of subextraction effects. *Cognitive Linguistics* 30: 719–50. [CrossRef]
Chaves, Rui P., and Jeruen E. Dery. 2019. Frequency effects in subject islands. *Journal of Linguistics* 55: 475–521. [CrossRef]
Chaves, Rui P., and Michael T. Putnam. 2020. *Unbounded Dependency Constructions: Theoretical and Experimental Perspectives*. Oxford: Oxford University Press. [CrossRef]
Chaves, Rui P., and Stephanie Richter. 2020. Investigating the role of verb frequency in factive and manner-of-speaking islands. In *Proceedings of the Cognitive Science Society (COGSCI)*. Austin: Cognitive Science Society, vol . 42, pp. 1771–77.
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Language*. Edited by Michael Kenstowicz. Cambridge: MIT Press, pp. 1–52.
Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Edited by Robert Freidin, Carlos P. Otero and Maria Luisa Zubizarreta. Cambridge and London: MIT Press, pp. 133–66.
Culicover, Peter W., Giuseppe Varaschin, and Susanne Winkler. 2022. The radical unacceptability hypothesis. *Languages* 7: 96. [CrossRef]
Culicover, Peter W., and Susanne Winkler. 2018. Freezing: Between grammar and processing. In *Freezing: Theoretical Approaches and eMpirical Domains*. Edited by Jutta M. Hartmann, Marion Jäger, Andreas Kehl, Andreas Konietzko and Susanne Winkler. Berlin and New York: Mouton de Gruyter, pp. 353–86. [CrossRef]
Culicover, Peter W., and Susanne Winkler. 2022. Parasitic gaps aren't parasitic, or, the case of the uninvited guest. *The Linguistic Review* 39: 1–35. [CrossRef]
den Dikken, Marcel. 2018. *Dependency and Directionality*. Cambridge: Cambridge University Press. [CrossRef]
Dowty, David R. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: Reidel. [CrossRef]
Featherston, Sam. 2008. The decathlon model of empirical syntax. In *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. Edited by Stephan Kepser and Marga Reis. Berlin: Mouton de Gruyter, pp. 187–208. [CrossRef]
Featherston, Sam. 2009. A scale for measuring well-formedness: Why syntax needs boiling and freezing points. In *The Fruits of Empirical Linguistics, Vol. 1: Process*. Edited by Sam Featherston and Susanne Winkler. Berlin: de Gruyter, pp. 47–74. [CrossRef]
Featherston, Sam. 2019. The decathlon model. In *Current Approaches to Syntax: A Comparative Handbook*. Edited by András Kertész, Edith Moravcsik and Csilla Rákosi. Berlin and Boston: Mouton de Gruyter, pp. 155–86. [CrossRef]
Featherston, Sam. 2020. Can we build a grammar on the basis of judgments? In *Linguistic Intuitions: Evidence and Method*. Edited by Samuel Schindler, Anna Drożdżowicz and Karen Brøcker. Oxford: Oxford University Press, pp. 165–88. [CrossRef]
Gerbrich, Hannah, Vivian Schreier, and Sam Featherston. 2019. Standard items for English judgment studies: Syntax and semantics. In *Experiments in Focus: Information Structure and Semantic Processing*. Edited by Sam Featherston, Robin Hörnig, Sophie von Wietersheim and Susanne Winkler . Boston & Berlin: de Gruyter, pp. 305–28. [CrossRef]
Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68: 1–76. [CrossRef]
Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, Language, Brain*. Edited by Alec Marantz, Yasushi Miyashita and Wayne O'Neil. Cambridge: MIT Press, pp. 95–126. [CrossRef]
Gibson, Edward, Steve Piantadosi, and Kristina Fedorenko. 2011. Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass* 5: 509–24. [CrossRef]
Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
Goldberg, Adele E. 2013. Backgrounded constituents cannot be "extracted". In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 221–38. [CrossRef]
Graf, Thomas. 2015. The syntactic algebra of adjuncts. In *Proceedings of the Chicago Linguistic Society (CLS)*. Chicago: Chicago Linguistic Society, vol . 49, pp. 101–13.
Greco, Ciro, Marco Marelli, and Liliane Haegeman. 2017. External syntax and the Cumulative Effect in subject sub-extraction: An experimental evaluation. *The Linguistic Review* 34: 479–531. [CrossRef]
Haegeman, Liliane, Ángel L. Jiménez-Fernández, and Andrew Radford. 2014. Deconstructing the subject condition in terms of cumulative constraint violation. *The Linguistic Review* 31: 73–150. [CrossRef]
Hofmeister, Philip, and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415. [CrossRef]
Hopper, Paul J., and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56: 251–99. [CrossRef]
Hu, Xuhui. 2018. *Encoding Events: Functional Structure and Variation*. Oxford: Oxford University Press.

Huang, Cheng-Teh James. 1982. Logical relations in Chinese and the Theory of Grammar. Ph. D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Jurka, Johannes. 2010. The Importance of Being a Complement: CED-Effects Revisited. Ph. D. thesis, University of Maryland, College Park, MD, USA.
Jurka, Johannes. 2013. Subject islands in German revisited. In *Experimental Syntax and Island Effects*. Edited by JJon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 265–85. [CrossRef]
Kehl, Andreas. 2021. Adjunct Islands in English: Theoretical Perspectives and Experimental Evidence. Ph. D. thesis, University of Tübingen, Tübingen, Germany.
Kohrt, Annika, Trey Sorensen, and Dustin A. Chacón. 2018. The real-time status of semantic exceptions to the adjunct island constraint. In *Proceedings of the Chicago Linguistic Society (CLS)*. Chicago: Chicago Linguistic Society, vol . 54, pp. 225–36.
Konietzko, Andreas. 2021. PP Extraction from Subject Islands in German. *Glossa*, accepted for publication.
Kuno, Susumu. 1987. *Functional Syntax: Anaphora, Discourse and Empathy*. Chicago and London: The University of Chicago Press.
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian wh-extraction. *Natural Language & Linguistic Theory* 36: 743–79. [CrossRef]
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language* 95: 393–420. [CrossRef]
Liu, Yingtong, Elodie Winckel, Anne Abeillé, Barbara Hemforth, and Edward Gibson. 2022. Structural, functional, and processing perspectives on linguistic island effects. *Annual Review of Linguistics* 8: 495–525. [CrossRef]
Müller, Christiane. 2019. *Permeable Islands: A Contrastive Study of Swedish and English Adjunct Clause Extractions*. Lund: Centre for Languages and Literature, Lund University.
Narita, Hiroki. 2014. *Endocentric Structuring of Projection-Free Syntax*. Amsterdam and Philadelphia: John Benjamins.
Ness, Tal, and Aya Meltzer-Asscher. 2019. When is the verb a potential gap site? The influence of filler maintenance on the active search for a gap. *Language, Cognition and Neuroscience* 34: 936–48. [CrossRef]
Phillips, Colin. 2013. On the nature of island constraints I: Language processing and reductionist accounts. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 64–108. [CrossRef]
Polinsky, Maria, Carlos G. Gallo, Peter Graff, Ekaterina Kravtchenko, Adam Milton Morgan, and Anne Sturgeon. 2013. Subject islands are different. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 286–309. [CrossRef]
Ramchand, Gillian Catriona. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press. [CrossRef]
Rapoport, Tova R. 1993. Stage and adjunct predicates: Licensing and structure in secondary predication constructions. In *Knowledge and Language, Vol. 2: Lexical and Conceptual Structure*. Edited by Eric Reuland and Werner Abraham. Dordrecht, Boston and London: Kluwer, pp. 157–82. [CrossRef]
Rapoport, Tova R. 2019. Secondary predication. In *The Oxford Handbook of Event Structure*. Edited by Robert Truswell. Oxford: Oxford University Press, pp. 426–55.
Roland, Douglas, Frederic Dick, and Jeffrey L. Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57: 348–79. [CrossRef] [PubMed]
Rothstein, Susan. 2004. *Structuring Events: A Study in the Semantics of Lexical Aspect*. Malden: Blackwell. [CrossRef]
Rothstein, Susan. 2017. Secondary predication. In *The Wiley Blackwell Companion to Syntax*, 2nd ed. Edited by Martin Everaert and Henk van Riemsdijk. Hoboken: Wiley, vol. 6, pp. 3872–901. [CrossRef]
Sag, Ivan A., Philip Hofmeister, and Neal Snider. 2008. Processing complexity in subjacency violations: The complex noun phrase constraint. In *Proceedings of the Chicago Linguistic Society (CLS)*. Chicago: Chicago Linguistic Society, vol . 43, pp. 215–29.
Santorini, Beatrice. 2019. (Un?)expected Movement. Available online: https://www.ling.upenn.edu/~beatrice/examples/movement.html (accessed on 27 March 2022).
Simpson, Jane. 2005. Depictives in English and Walpiri. In *Secondary Predication and Adverbial Modification: The Typology of Depictives*. Edited by Nikolaus P. Himmelmann and Eva Schultze-Berndt. Oxford: Oxford University Press, pp. 69–106. [CrossRef]
Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43: 155–67. [CrossRef]
Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34: 307–44. [CrossRef]
Sprouse, Jon, and Norbert Hornstein. 2013. Experimental syntax and island effects: Toward a comprehensive theory of islands. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 1–17. [CrossRef]
Sprouse, Jon, Matthew W. Wagers, and Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88: 82–123. [CrossRef]
Sprouse, Jon, Matthew W. Wagers, and Colin Phillips. 2013. Deriving competing predictions from grammatical approaches and reductionist approaches to island effects. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 21–41. [CrossRef]
Staub, Adrian. 2007. The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology* 33: 550–69. [CrossRef]

Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax and Semantics* 10: 80–126. [CrossRef]
Tenny, Carol. 1995. How motion verbs are special: The interaction of semantic and pragmatic information in aspectual verb meanings. *Pragmatics & Cognition* 3: 31–73. [CrossRef]
Truswell, Robert. 2007. Extraction from adjuncts and the structure of events. *Lingua* 117: 1355–77. [CrossRef]
Truswell, Robert. 2011. *Events, Phrases, and Questions*. Oxford: Oxford University Press. [CrossRef]
Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66: 143–60. [CrossRef]
Wagers, Matthew W. 2013. Memory mechanisms for wh-dependency formation and their implications for islandhood. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 161–85. [CrossRef]
Winkler, Susanne. 1997. *Focus and Secondary Predication*. Berlin and New York: Mouton de Gruyter. [CrossRef]

languages

MDPI

Article

On the Nature of Syntactic Satiation

William Snyder

Department of Linguistics, University of Connecticut, Storrs, CT 06269-1145, USA; william.snyder@uconn.edu

Abstract: In syntactic satiation, a linguist initially judges a sentence type to be unacceptable but begins to accept it after judging multiple examples over time. When William Snyder first brought this phenomenon to the attention of linguists, he proposed satiation as a data source for linguistic theory and showed it can be induced experimentally. Here, three new studies indicate (i) satiation is restricted to a small, stable set of sentence types; (ii) after satiation on one sentence type (e.g., *wh*-movement across … *wonder whether* … or … *believe the claim* …), acceptability sometimes increases for distinct but syntactically related sentence types (… *wonder why* …; … *accept the idea* …); (iii) for sentence types susceptible to satiation, the difficulty of inducing it (e.g., number of exposures required) varies systematically; and (iv) much as satiation in linguists persists over time, experimentally induced satiation can persist for at least four weeks. These findings suggest a role for satiation in determining whether the perceived unacceptability of two sentence types has a common source.

Keywords: syntactic satiation; linguistic judgments; island effects; experimental syntax

Citation: Snyder, William. 2022. On the Nature of Syntactic Satiation. *Languages* 7: 38. https://doi.org/10.3390/languages7010038

Academic Editors: Anne Mette Nyvad and Ken Ramshøj Christensen

Received: 9 August 2021
Accepted: 1 February 2022
Published: 17 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

1.1. Overview of the Project

In generative linguistics, information about a person's mental grammar comes primarily from that person's judgments of acceptability: certain combinations of form and meaning are fully acceptable, while others are not. The standard idealization is that any given native-speaker consultant who is asked, on different occasions, to judge the same <form, meaning> pair will provide the same judgment on each occasion.

A systematic exception is presented by "satiation" effects: for certain initially unacceptable sentence types, after a linguist has judged multiple examples over a period of time, the perceived acceptability increases. Satiation calls out for investigation, not only because linguistic theories need to take account of its possible effects on the data they use but also because it may provide new insights into the basic phenomena that linguistic theories are meant to explain.

This article performs some of the necessary groundwork for linguistic investigation of satiation by providing evidence for the following points:

(1) a. While satiation effects were first noticed informally among professional linguists, they can also be induced in non-linguists, under controlled conditions in the laboratory;
b. Satiation effects induced in the laboratory are replicable, in the sense that the set of sentence types that potentially satiate is consistent across studies (and for the majority of sentence types, satiation does not occur);
c. Satiation effects for different types of "satiable" grammatical violation have different signatures (e.g., in the number of exposures typically needed before satiation occurs and in the typical percentage of experimental participants whose judgment changes).

The objective will be to show that investigation of satiation can broaden the range of empirical phenomena (and, thus, sources of data) bearing on key linguistic issues, including

in particular a whole range of issues concerning the nature and status of acceptability judgments.

1.2. Overview of Satiation

Building on earlier unpublished work by Karin Stromswold, Snyder (2000) published a squib drawing linguists' attention to the phenomenon Stromswold had termed "syntactic satiation". For a highly circumscribed set of sentence types, linguists sometimes experience a "shift" in their native-speaker acceptability judgments. The paradigm case is (2).

(2) Who does John wonder whether Mary likes?
(*Answer: He wonders whether she likes Pat*.)

On first exposure to an example like (2), a linguist may have thought it sounded starkly unacceptable. Yet, by the time the linguist was teaching an introductory course on syntax and presented an example like (2) to the students, the perception of grammatical impossibility may have been weaker, or even absent altogether. If so, the linguist had experienced satiation on that sentence type.

The description in (Snyder 2000), together with anecdotal reports and personal experience, motivates the characterization in (3).

(3) Characteristics of syntactic satiation when experienced by linguists:
   a. Lexical Generality: Satiation operates at the level of a grammatical structure. The increased acceptability of the structure is general, extending beyond the specific sentences that caused satiation—at a minimum, to sentences with different open-class lexical items.
   b. Structural Specificity: Only a limited number of sentence structures (i.e., types of grammatical violation) are potentially affected by satiation.
   c. Between-speaker Consistency: At least across native speakers of English, the same sentence types (notably sentences involving *wh*-extraction of an argument from a *Wh*-Island, Complex NP, or Subject Island) are the ones that are, at least in principle, susceptible to satiation.
   d. Within-speaker Persistence: Once an individual has experienced satiation on a given sentence type, the increased acceptance persists for a considerable period of time, even in the absence of routine exposure to sentences of that type.

In judging whether an experimental effect qualifies as "satiation" in the relevant sense, the characteristics in (3) will serve as a guide.

In the sections that follow, three new experimental studies are presented and discussed in light of the following questions:

(4) a. Can satiation in fact be demonstrated in the laboratory, or were the findings in (Snyder 2000) simply an experimental artifact, as proposed in Sprouse's (2009) Response Equalization Hypothesis (REH)? (Sections 2 and 3)
   b. When satiation is induced in the laboratory, does it persist beyond the experimental session? (Section 4)
   c. Does the difficulty of inducing satiation vary between different sentence types that are susceptible to the effect? (Section 5)
   d. Does satiation on one sentence type ever carry over to judgments of related sentence types, for example, from *Whether*-Island violations to sentences violating another type of *wh*-island? (Section 6)
   e. How sensitive is the satiation phenomenon to details of experimental methodology? What aspects of the methodology appear to matter? (Section 7)

Section 7 includes a survey of the literature on satiation. Section 8 turns to larger questions: the nature of satiation and its relevance to the objectives of generative linguistics.

2. Review of the Original Findings

An important component of (Snyder 2000) was the evidence suggesting satiation can be induced in the laboratory, under controlled experimental conditions, and measured

objectively.[1] Yet, Sprouse (2009) raised an important concern: the findings might have been due to what he termed "response equalization", rather than syntactic satiation. To facilitate discussion of the issue, this section will describe Snyder's (2000) experiment in detail. Section 3 will present a new experiment based on that study but modified to preclude response equalization.

2.1. Overview of Methodology and Findings

The experimental task in (Snyder 2000) took the form of a lengthy, printed questionnaire. Native English speakers, with no prior exposure to linguistics, were required to provide acceptability judgments on sentence-meaning pairs. A certain number of (initially) unacceptable sentence structures were systematically repeated in the course of the questionnaire. Thus, the participant received a compressed version of the linguist's experience of judging structurally equivalent sentences on multiple, distinct occasions.

On each page there was a single item like (5) (Snyder 2000, p. 576).

(5) (Context: Maria believes the claim that Beth found a $50 bill.) Test Sentence: "What does Maria believe the claim that Beth found?" Judgment: ____ (Y/N)

Prior to starting, participants were told they would be asked for a series of 60 judgments. On each page there would be a declarative sentence (the "context") and then an interrogative sentence (the "test sentence"). Participants were instructed to provide a Yes/No judgment: Is the test sentence grammatically possible in English, given the meaning that fits the context? In other words, could the test sentence have the intended meaning and still be accepted as "English", in their personal opinion? Participants were advised that many items would be similar to one another, but they should not look back to previous pages or try to remember previous answers. Given that no two items would be identical, and given that the differences might be important, they should simply provide an independent judgment on each new test sentence and then move on.

Fifty of the items corresponded to a series of five experimental blocks (although this structure was invisible to participants). Each block contained items of the following types, in pseudo-random order: three fully grammatical items and seven items that would typically be perceived as anywhere from mildly to severely unacceptable, namely one item of each type in (6).[2]

(6) a. Adjunct-Island violation
(Context: Paula wrote two novels before meeting the great playwright.)
Test Sentence: "Who did Paula write two novels before meeting?"

b. Complex Noun Phrase Constraint (CNPC) violation
(Context: John believes the claim that Mary likes Tony.)
Test Sentence: "Who does John believe the claim that Mary likes?"

c. Left Branch Constraint (LBC) violation
(Context: Bill knows that Alice smoked two cigarettes.)
Test Sentence: "How many does Bill know that Alice smoked cigarettes?"

d. Subject-Island violation
(Context: Sally knows that a bottle of vinegar fell on the floor.)
Test Sentence: "What does Sally know that a bottle of fell on the floor?"

e. *That*-trace violation
(Context: Fred believes that Greta frightened Bob.)
Test Sentence: "Who does Fred believe that frightened Bob?"

f. *Want-for* violation
(Context: Bob wants Vanessa to buy a hammer.)
Test Sentence: "What does Bob want for Vanessa to buy?"

g. Whether-Island violation
(Context: Dmitri wonders whether John drinks coffee.)
Test Sentence: "What does Dmitri wonder whether John drinks?"

In addition to the 50 test items, the experimental materials included six practice items immediately prior to Block 1 and four post-test items immediately following Block 5. (The distinction between these items and the actual test items was invisible to participants.) No two test items were ever identical: even within a single sentence type, almost all the open-class lexical items differed across sentences. There were two exceptions: CNPC violations in the body of the experiment—but not the post-test—consistently used the phrase *believe the claim*, and *Whether*-Island violations in the body of the experiment—but not the post-test—consistently used the phrase *wonder whether*.

An informal poll of linguists (all of them native speakers of English) indicated to Snyder that the phenomenon of syntactic satiation was relevant (at least) to *wh*-extraction of an argument across a *Whether* Island (6g) and out of a complex noun phrase of the type in (6b). In contrast, there appeared to be no satiation on LBC violations (6c) or *That-trace* violations (6e).[3] Thus, Snyder reasoned that if syntactic satiation could indeed be induced by his task, there ought to be a systematic tendency for participants to become more accepting of *Whether*-Island violations and/or CNPC violations by the end of the experiment. There should not, however, be increased acceptance of LBC or *That-trace* violations. (For the other sentence types in (6), the possibility of satiation was treated as an open question.)

The findings were as follows: As predicted, for both *Whether* and CNPC items there was a significant increase in acceptance from the beginning (Blocks 1 and 2) to the end (Blocks 4 and 5) of the questionnaire (two-tailed $p < .05$ by Binomial Test).[4] In contrast, for LBC and *That-trace*, there was no appreciable change. Hence, the findings were broadly consistent with the possibility that the task was inducing the same kind of judgment change that linguists sometimes experience. (Of the other sentence types, only Subject Islands showed any appreciable increase, and it was only marginally significant; $p < .07$.)

The four post-test items following Block 5 were two fully grammatical fillers, plus the two items in (7).

(7) a. Complex Noun Phrase Constraint (CNPC) violation, with *accept the idea*
(Context: Madge accepted the idea that Bob would run for mayor.)
Test Sentence: "What did Madge accept the idea that Bob would do?"

b. Whether-Island violation, with *ask whether*
(Context: Mildred asked whether Ted had visited Stonehenge.)
Test Sentence: "What did Mildred ask whether Ted had visited?"

To check for a possible "carryover" effect from judging CNPC violations with *believe the claim* (as in Blocks 1–5; cf. 6b), to *accept the idea* (7a), Snyder focused on participants who had initially rejected both of the CNPC violations in Blocks 1 and 2. (The intention was to focus on individuals whose grammar had clearly excluded such sentences prior to the experiment.) These participants were first classified as satiating or not satiating on *believe the claim*, based on whether they accepted at least one of the two *'believe the claim'* items in Blocks 4 and 5. They were then cross-classified as accepting or rejecting the post-test item, (7a).

Both in (Snyder 2000) and in the new experiments reported below, a participant is classified as having "satiated" on a given sentence type if (and only if) one of the following three situations holds true: (i) the exemplars in the first two blocks of the study were both rejected and exactly one of the exemplars in the final two blocks was accepted, (ii) the exemplars in the first two blocks were both rejected and the exemplars in the final two blocks were both accepted, or (iii) exactly one of the exemplars in the first two blocks was accepted and both of the exemplars in the final two blocks were accepted.

The rate of acceptance of the post-test item among participants who had consistently rejected the CNPC violations, both in Blocks 1 and 2 and in Blocks 4 and 5, was calculated as a baseline. A binomial test was then used to assess the data from participants who had likewise rejected the CNPC violations in Blocks 1 and 2 but accepted at least one of the CNPC violations in Blocks 4 and 5 (i.e., had satiated), in order to answer the following question: What was the probability of obtaining, simply by chance, an acceptance rate for the post-test item that was as high as (or even higher than) the rate observed in these latter

participants? "Simply by chance" meant that the probability was calculated under the null hypothesis that, in general (among participants who rejected both items in Blocks 1 and 2), the participants who satiated (i.e., accepted at least one of the items in Blocks 4 and 5) had the same probability of accepting the post-test item as the participants who had not satiated.

In Snyder's data, all 22 participants had rejected the CNPC violations (i.e., with *believe the claim*) in Blocks 1 and 2. Of those 22, 17 also rejected the CNPC violations in Blocks 4 and 5. Only four of these 17 "non-satiators" accepted the post-test item with *accept the idea*. In contrast, among the five satiators, four accepted the post-test item. Under the null hypothesis that the general acceptance rate for satiators was the same as for non-satiators, namely 4/17 = 23.5%, the likelihood of seeing acceptance by at least four out of five satiators simply by chance is given by the binomial test: in the present case, two-tailed $p < .05$. Hence, there was significant carryover.

In the case of *Whether*-Island violations, 18 of the 22 participants rejected the items (i.e., with *wonder whether*) in Blocks 1 and 2. Of these 18, seven also rejected the *wonder-whether* items in Blocks 4 and 5. Only three of these non-satiators accepted the post-test item (with *ask whether*). This provided a baseline acceptance rate of 3/7 = 42.9%. Of the 11 satiators, however, 10 accepted the post-test item (binomial $p < .005$). Hence, there was also significant carryover for *Whether* Islands.[5]

In sum, Snyder (2000) obtained statistically reliable satiation on argument *wh*-extraction from both the complex-NP (*believe the claim*) environment and the w*onder-whether* environment, although far fewer participants showed the effect with complex NPs (five out of 22, as opposed to 11 of 22 for *whether*). Moreover, the satiation overwhelmingly "carried over" from *believe the claim* to *accept the idea* and from *wonder whether* to *ask whether*: four of the five satiators on CNPC violations exhibited carryover, as did 10 of the 11 satiators on *Whether* Islands.

2.2. Some Possible Concerns

A few further details of methodology are important for the present discussion. A major issue in any type of work with acceptability judgments is the fact that many different factors can influence them. These include not only the grammatical structure of the sentence being judged, but also the choices of open-class lexical items and the characteristics of the test item that was judged immediately prior. Therefore, alongside satiation, one of the possible reasons for participants to become more accepting of a given sentence type, as they work their way through an experiment, is that the specific examples presented later in the experiment are somehow intrinsically more acceptable, for reasons independent of their grammatical structure (e.g., due to the open-class lexical items that they happen to contain). Another possibility is that the specific test items positioned immediately prior to the sentences of interest made the earlier sentences seem less acceptable, and/or the later ones seem more acceptable, than they would ordinarily.

A simple way to minimize these possibilities is to counterbalance, across participants, the order of presentation: half the participants receive the items in forward order, and the other half receive the same items but in reverse order. Snyder (2000) therefore gave half of his participants a questionnaire containing the 50 test items in the order "..., Item 1, Item 2, Item 3, ..." and gave the other half the same items but in the order "..., Item 50, Item 49, Item 48, ...". Any items that were intrinsically more acceptable than others of the same type would yield an increase in acceptability for half the participants but an equally strong decrease for the other half. Similarly, if judging a certain test item had a special effect on the participant's next judgment, then this effect would apply to different "next judgments" in the different orders of presentation. Crucially, if the experiment induced actual satiation on sentences of a given grammatical type, then it should yield increasing acceptance not only overall but also both in the subset of participants who received a "forward" order of presentation and in the subset who judged the same items but in reverse order.[6]

2.3. Response Equalization?

Let's now consider Sprouse's (2009) Response Equalization Hypothesis (REH). One type of task effect that is not addressed simply by counterbalancing the order of presentation, and that must be addressed separately, is the following. Suppose that participants come to any Yes/No task, such as the one in (Snyder 2000), with an expectation that exactly half the test items will have an expected answer of "Yes". A problem, then, is that, for 50 of the 60 items in Snyder's experiment (i.e., the five blocks of ten mentioned above), there was a ratio of seven items with a grammatical violation for every three items that were fully grammatical. Now, blocking of the items was invisible to the participants, and exactly half of the other 10 items (i.e., practice and post-test items) were fully grammatical. Therefore, participants saw a single series of 60 items, and 40 of them (66.7%) contained a grammatical violation.

Assuming participants noticed the discrepancy between the expected frequency of "Yes" items (50.0%) and the actual frequency (presumably 33.3%, for an unchanging native-speaker grammar of English), the REH says participants should have become more willing to say "Yes" as the experiment progressed. To make sure that Snyder's (2000) findings were not simply due to response equalization, the best approach is to rerun the experiment with exactly one change: add enough fully grammatical items for a 1:1 balance. This will be the first of three new experiments reported below.[7]

3. Experiment I: A Direct Test of the Response Equalization Hypothesis

3.1. Materials

Experiment I was identical to the experiment in (Snyder 2000) except that 20 new, fully grammatical test items were added to the questionnaire, so as to create a perfect balance: 40 items that were fully grammatical and 40 that violated a grammatical constraint. For each of the experimental blocks in Version A, four of the new items were randomly selected and inserted among the original 10 items, as follows: 1 2_3 4_5 6_7 8_9 10. Following these additions, each of the five blocks contained seven fully grammatical items and seven grammatical violations (one item for each of the seven types in (6), above), and there were never more than two expected "NO" items in a row. A new Version B was created from Version A by reversing the order of the resulting 70 test items. Together with the six practice items and four post-test items, this yielded 80 items per participant.

In keeping with the original materials of (Snyder 2000), the new grammatical items were designed to be comparable in their structural complexity to the ungrammatical items. Some representative examples are provided in (8b,d,f).

(8) a. CNPC violation, with *believe the claim*:
(Context: Maria believes the claim that Beth found a $50 bill.)
Test Sentence: "What does Maria believe the claim that Beth found?"

b. Grammatical item, with *claim to believe*:
(Context: John claims to believe that Mary likes Tony.)
Test Sentence: "Who does John claim to believe that Mary likes?"

c. Whether-Island Violation, with *wonder whether*:
(Context: Henry wonders whether George discovered the answer.)
Test Sentence: "What does Henry wonder whether George discovered?"

d. Grammatical item, with *wonder what*:
(Context: Gina wonders whether Einstein discovered relativity.)
Test Sentence: "Who wonders what Einstein discovered?"

e. LBC violation, with *how many ... books*:
(Context: Edwin thinks Margaret read three books.)
Test Sentence: "How many does Edwin think Margaret read books?"

f. Grammatical item, with *how many books*:
(Context: Edward thinks that Anne read ten books.)
Test Sentence: "How many books did Edward think that Anne had read?"

Under the REH account of Snyder's (2000) findings, the prediction for Experiment I (where participants now see the same number of expected "YES" and expected "NO" items) is that there will be no systematic tendency for any sentence type to be accepted more often at the end than at the beginning of the experiment. In contrast, what we might call the "Satiation Hypothesis" predicts an increased likelihood of "Yes" responses at later points in the experiment for *Whether*-Island and/or CNPC violations but no systematic tendency toward increasing acceptance of *That*-trace or LBC violations.

3.2. Plan for Data Analysis

In a yes–no task, the responses cannot be expected to obey a normal (Gaussian) distribution. Snyder (2000) therefore relied primarily on binomial tests and Fisher Exact Tests, which are both "non-distributional" in the sense of not assuming a normal distribution. Here, the approach to data analysis will again rely on non-distributional methods of two main types. First, for each of the initially unacceptable sentence types, a Wilcoxon Signed-Rank test will be used to assess whether "Yes" responses were significantly more frequent at the end of the experiment (in the final two blocks) than at the beginning (in the initial two blocks).

Second, whenever possible, mixed-effect (ME) logistic regression will be used as a follow-up test. The logistic extension to linear regression is in many ways ideal for the analysis of yes–no judgment data, but sometimes, there are difficulties in achieving convergence (i.e., in fitting a model to the dataset), especially if the number of participants is relatively small. Given that convergence is not always possible, the role of ME Logistic Regression will be secondary: in the event that convergence cannot be achieved, the results of the Wilcoxon Tests will have to suffice. (In practice, such a situation will arise during the analysis of data from Experiment II, below.)

When applying ME logistic regression, the search for a model fit will always begin with a "maximally" specified model (cf. Barr et al. 2013), which will then be simplified if necessary in order to achieve convergence. A limit on simplification, however, will be that the Random Effects (RE) portion of the model must always include "random intercepts" for individual participants and for individual test items and must include by-participant "random slopes" for the effect of each of the major factors in the experiment. (For the present purposes, the major factors are the sentence type being judged and the block of the experiment in which the judgment is made.) This ensures that the model is appropriately adjusted for (i) variation in the overall willingness of a participant to say "yes" to test items in general (i.e., the by-participant random intercept), (ii) the participant's general willingness to say "yes" to each of the different types of sentence (i.e., the by-participant random slopes for Type), and (iii) the participant's general willingness to say "yes" in each successive Block of the experiment (i.e., the by-participant random slope for Block). It also ensures that the model adjusts for variation across the different sentences (i.e., "ItemCodes") that exemplify a particular sentence type (i.e., within any single experimental treatment).

A small change from (Snyder 2000) is that the blocks of Experiment I (as well as Experiments II and III below) will be numbered from 0 to 4, rather than 1–5. This has the desirable consequence that, for each of the (initially) unacceptable sentence types, the block number can be interpreted as the participant's number of previous exposures to that sentence type during the experiment.

One special strength of ME Logistic Regression is its ability to evaluate a given participant's response to a test sentence relative to that same participant's responses to control sentences. The control sentences (in all of Experiments I–III) will be grammatically well-formed sentences that are similar to the test sentences in their structural complexity and that are judged in the same block as the corresponding test item. If participants experience genuine satiation on sentences of type T, then we expect ME logistic regression to reveal a significant interaction between block number and sentence type, for sentence type T.

More precisely, ME logistic regression will be conducted with one level of each factor specified as a baseline for use in "treatment contrasts" (i.e., pairwise comparisons) with

each of the other levels of that factor. For Type, the baseline level will be "Good" (i.e., within each block, the results for the seven fully grammatical items). A treatment contrast will then be calculated for each of the seven other (i.e., deviant) sentence types. Crucially, for each of these non-baseline levels, the analysis will check for an interaction effect: did the effect of "changing" from the grammatical items (the baseline) to an item of that type differ significantly, as a function of the experimental block in which the judgments were made?

Finally, evidence of increased acceptance at the end of the experiment (in the form of a significant Wilcoxon test and, when ME logistic regression converges, a significant interaction effect) is necessary, but not sufficient, for a claim of satiation on T. If participants exhibit genuine satiation of the kind characterized earlier in (3), then we expect some additional findings, and we need to confirm their presence. Specifically, the increased acceptability of a given sentence type following satiation should be evident regardless of the order in which sentences were presented. Hence, the next step will be to examine the data from Versions A and B separately. If genuine satiation occurred, we expect each version to show a statistically significant increase, from the beginning to the end of the experiment, in the frequency of acceptance.

3.3. Experimental Participants and Procedure

The participants in Experiment I were 22 undergraduate students, all native speakers of English, who were recruited by means of printed flyers posted on campus. Compensation was provided in the form of a $5 gift card, redeemable at the university bookstore. Participants were brought into an individual testing room and told the instructions (which were also provided in printed form). Participants then received the materials in the form of a printed booklet, exactly as in (Snyder 2000). Completion of the task took about 15 min.

3.4. Checking for Outliers

Prior to running inferential statistics, the data were checked for participants more than two standard deviations from the group average on either expected "YES" or expected "NO" items, because any such participants may not have understood the instructions. Indeed, two participants were more than two standard deviations below the group mean on acceptance of expected "YES" items and were excluded from further analysis, leaving $N = 20$.[8]

3.5. Primary Analysis: Wilcoxon Tests

Wilcoxon Signed-Rank tests were used to assess statistical reliability of changes in acceptance rate, for each sentence type, between the first two blocks (0 and 1) and the final two blocks (3 and 4). The main results were as follows. Acceptance in Blocks 3 and 4 was significantly greater for *whether* items ($W = -93$, $n_{s/r} = 14$, $Z = -2.9$, $p < .005$), but there was no significant change for any other sentence type (all $p > .10$). The data are shown graphically in Figures 1–4.[9]

For *whether* items, when the 20 participants are viewed individually, some 14 showed a change between the initial two blocks and the final two, and in 13 cases, it was an increase. (Nine increased from 0/2 to 1/2, three increased from 0/2 to 2/2, and one increased from 1/2 to 2/2. The individual showing a decrease changed from 2/2 to 1/2.) Among the six participants whose level of acceptance was unchanged, three consistently rejected the sentences, and three consistently accepted them.

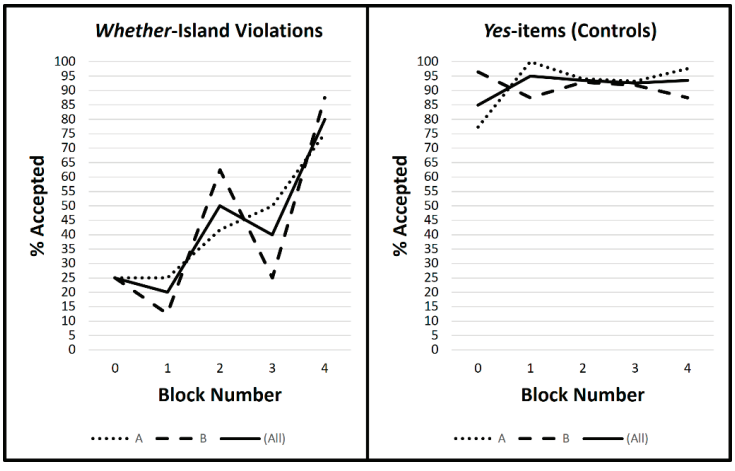


Figure 1. (**Left**) Percentage of participants in each block of Experiment I, who accepted the *Whether*-Island violation; Version A (N = 12) used forward presentation; Version B (N = 8) used reverse order; "All" indicates the total (N = 20). (**Right**) Mean percentage of the "Yes" items that were accepted; each participant judged seven items per block.

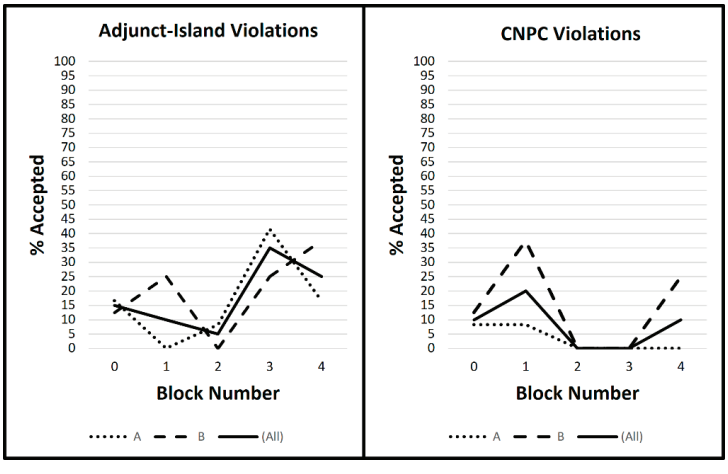


Figure 2. Experiment I, Adjunct-Island and CNPC violations.

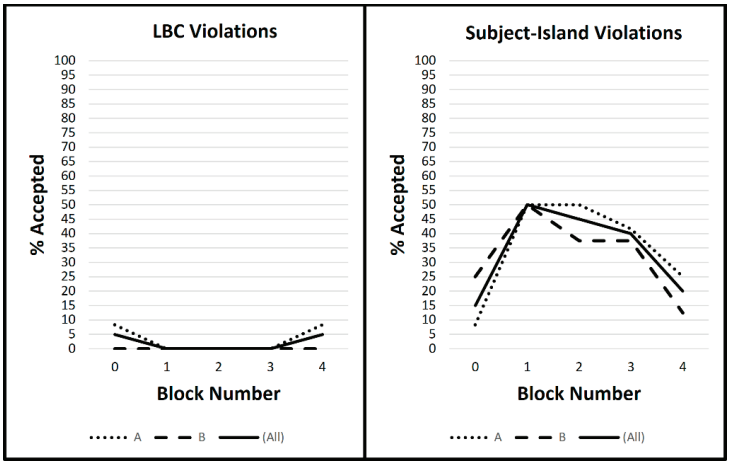


Figure 3. Experiment I, LBC and Subject-Island violations.

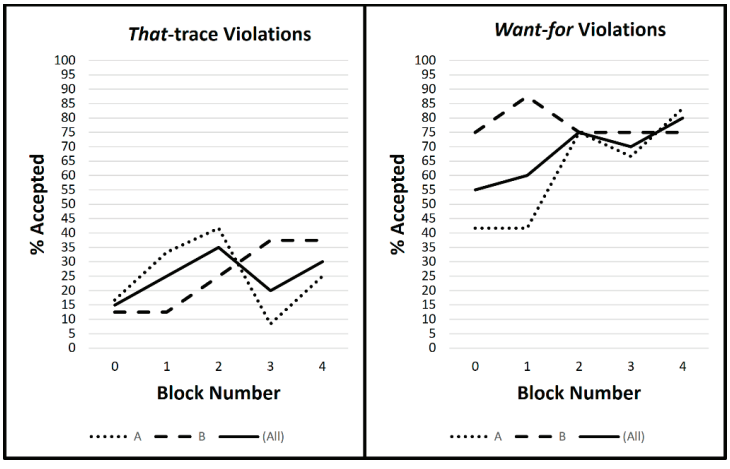


Figure 4. Experiment I, *That*-trace and *Want-for* violations.

3.6. Cross-Checking: ME Logistic Regression

Linear modeling was performed using R (version 3.2.3; R Core Team 2015) and the lme4 software package (version 1.1–11; Bates et al. 2015). An ME logistic model was constructed using lme4's "glmer" function. In the notation of the lme4 package, the model was specified as in (9).

(9) Response~Block*Type + (1 + Block + Type | Participant) + (1 | ItemCode) + (1 | Version)

Thus, the software searched for an optimally specified logistic-regression model with which to "predict" each participant's yes/no response to each test item, based on (i) the (integer) number of the block (0–4) in which the test item appeared, (ii) the grammatical type 'T' of the test item, and (possibly) (iii) an interaction effect between Block and Type for each (non-baseline) value of Type. As noted above, a significant interaction is precisely what we expect to see if participants experience satiation on a given sentence type.

The initial attempt to fit a model with structure (9) to the data from Experiment I was unsuccessful: the glmer program failed to converge. Inspection of the program's best attempt revealed two issues. First, in the RE structure, the random intercepts by Version

explained none of the variation in the dataset.[10] Second, in the fixed-effects structure for the program's "best attempt" at a fit, the main effect of Block had an estimated coefficient (0.16) that was an order of magnitude smaller than the coefficients for the main effects of the different levels of Type (which ranged from 1.73 to 8.21). A difference in scale of one or more orders of magnitude can prevent convergence. Hence, two changes were made. First, Version was removed from the RE structure in (9). Second, the factor Block was re-scaled: the values of Block in the dataset were simply divided by 10 (so that Block ranged from 0.0 to 0.4).

Following these changes, the program converged on the model summarized in Table 1.[11] As expected, pairwise comparisons showed that each of the ungrammatical levels of Type differed significantly from the grammatical (baseline) items. There was no main effect of Block ($p > .10$), and there was exactly one significant interaction of Block with Type, namely for Type = *Whether*; acceptance of *Whether* items increased significantly, as the participants progressed from Block 0 to Block 4.

Table 1. Table of fixed effects for Experiment I.

Predictor	Estimate	SE	*Z*	*p*
(Intercept)	2.90	0.44	6.64	<.001
Block	1.41	1.49	0.95	(>.10)
TypeAdjunct	−6.28	1.05	−5.98	<.001
TypeCNPC	−5.83	1.08	−5.41	<.001
TypeLBV	−10.64	3.40	−3.13	<.01
TypeSubject	−3.90	0.83	−4.68	<.001
TypeThat	−5.01	0.93	−5.38	<.001
TypeWant	−2.17	1.06	−2.04	<.05
TypeWhether	−5.68	1.03	−5.50	<.001
Block:TypeAdjunct	3.58	2.74	1.31	(>.10)
Block:TypeCNPC	−3.46	3.31	−1.05	(>.10)
Block:TypeLBV	0.18	5.99	0.03	(>.10)
Block:TypeSubject	−1.62	2.38	−0.68	(>.10)
Block:TypeThat	0.65	2.48	0.26	(>.10)
Block:TypeWant	3.11	2.74	1.14	(>.10)
Block:TypeWhether	9.58	3.00	3.20	<.01

Thus, the results of ME logistic regression are entirely consistent with the results from Wilcoxon tests: in Experiment I there was possible satiation on items with *wonder whether*, but (in contrast to Snyder 2000) there was no satiation on the complex-NP items with *believe the claim* ($p > .10$). Consistent with Snyder 2000, there was no satiation on any of the other sentence types tested.

3.7. Follow-Up Testing

The next question is whether the apparent satiation on *wonder whether* meets the additional criterion discussed above: Did acceptance increase, from the beginning to the end of the questionnaire, in both versions? Indeed, Versions A and B each showed the same general pattern as the full study. Overall (as noted above), fourteen participants showed a change, and in 13 cases, it was an increase. On Version A, seven participants changed, and in all cases, it was an increase. On Version B, seven participants changed, and in six cases, it was an increase. Hence, the findings conform very closely to what is expected in satiation.

Experiment I shows that satiation can indeed be obtained under laboratory conditions, at least for *wonder whether* items, even if participants judge a perfect balance of fully acceptable, versus initially unacceptable, sentences. The main difference from Snyder 2000 is the absence of a change for CNPC items. In Section 5, we will see evidence that the final sample size ($N = 20$) in Experiment I was far too low for reliable detection of satiation on CNPC sentences, but regardless, the specific sentence type (*wonder whether*) that showed satiation was also one of the types showing it in (Snyder 2000). Hence, the findings from Experiment I are fully in-line with Between-speaker Consistency (3c) (as well as Generality and Structural Specificity). Next, we check for Within-speaker Persistence (3d).

4. Experiment II: Persistence

Did the increase in acceptance of *Whether*-Island violations observed in Experiment I persist beyond the time of the experiment? To find out, each participant was invited to return for testing one month later. Of the 20 participants whose data were included in the analyses for Experiment I, 15 agreed to return.

Each of these participants was tested again, 4 to 5 weeks later, in much the same way as the first time. In almost all cases, if a participant (for example) received Version A at Time 1, then Version B was given at Time 2. One participant was accidentally given B at both Time 1 and Time 2. Among the other 14, eight received version A and six version B at Time 1; hence, six (of these 14) received A and eight received B at Time 2.

The predictions were as follows. If the satiation on *Whether* items in Experiment I quickly faded, then there should be no significant difference between participants' judgments at the beginning of Experiment I (Blocks 0 and 1), and the same participants' judgments one month later at the beginning of Experiment II (Blocks 0 and 1). In contrast, if the satiation that was detected on *Whether* items showed Within-speaker Persistence, then, at least for *Whether* items, the frequency of acceptance at the beginning of Experiment II should be significantly higher.

Moreover, if the satiation on *Whether* items persisted, this should be evident when we examine the data by participant. For example, someone who accepted neither of the examples in Blocks 0 and 1 of Experiment I but accepted one of the examples in Blocks 3 and 4 of Experiment I would be expected to accept at least one of the two examples in Blocks 0 and 1 of Experiment II.

4.1. Primary Analysis: Wilcoxon Tests

Wilcoxon Signed-Rank Tests were performed to check for increased acceptance at Time 2. For each sentence type, each participant's responses in Blocks 0 and 1 of Experiment I were compared against Blocks 0 and 1 of Experiment II. As predicted by Within-speaker Persistence, there was a significant increase for *Whether* items from the beginning of Experiment I ("Time 1") to the beginning of Experiment II ("Time 2"; $W = -45$, $n_{s/r} = 9$, $p < .01$). No other sentence type showed a significant increase. On *Whether*, when the participants are viewed individually, six were consistent across Times 1 and 2, and nine showed a change. In all cases, if there was a change, it was an increase: for four participants, from 0/2 "yes" responses at Time 1 to 1/2 "yes" at Time 2 and, for five participants, from 0/2 at Time 1 to 2/2 at Time 2. The full results are shown graphically in Figure 5.

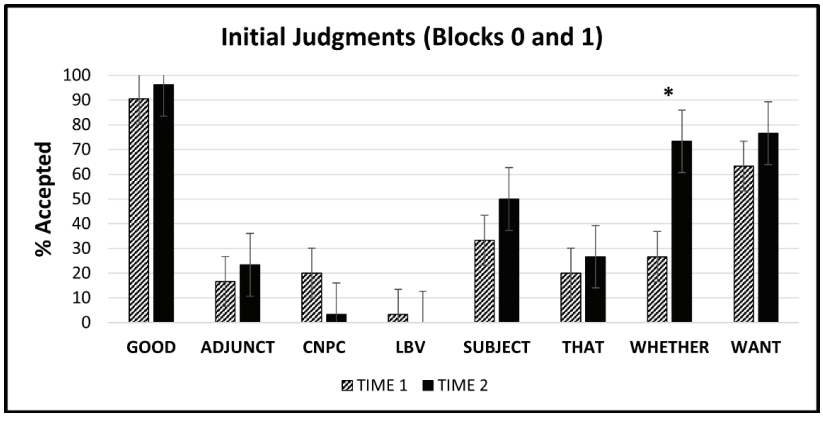


Figure 5. Experiment II: Acceptance of initial sentences (Blocks 0 and 1) at Times 1 and 2; Error bars show standard error; "*" indicates significance ($p < .05$) by Wilcoxon Signed-Rank Test.

The by-participant results are shown in Figure 6. Of the 15 individuals who participated in Experiment II, 11 rejected both of the *Whether*-Island violations at the beginning (Blocks 0 and 1) of Experiment I, and four accepted both of them. Of the 11 who rejected them, one participant continued to reject them at the end (Blocks 3 and 4) of Experiment I, while the other ten accepted at least one (i.e., they had satiated). As can be seen in the table, eight of the ten satiators accepted at least one of the two exemplars of a *Whether*-Island violation in Blocks 0 and 1 of Experiment II; the remaining two satiators both accepted one fewer than they had in Blocks 3 and 4 of Experiment I. (The four participants who accepted the exemplars in Blocks 0 and 1 of Experiment I continued to accept (in all but one case) the exemplars at the end of Experiment I and beginning of Experiment II.)

	Initially 0											*Initially 2*			
Participant ID:	**16**	**4**	**7**	**22**	**3**	**6**	**11**	**1**	**5**	**8**	**10**	**12**	**9**	**15**	**21**
T1,Beginning:	0	0	0	0	0	0	0	0	0	0	0	2	2	2	2
T1,End:	0	1	1	1	1	1	1	2	2	1	2	1	2	2	2
T2,Beginning:	0	1	1	1	2	2	2	2	2	0	1	2	2	2	2

T1 Satiation

T2 Maintenance of Satiation

Figure 6. By-participant findings in Experiment II, showing the number (0–2) of Whether-Island violations accepted at three time points (Beginning and End of Experiment I and Beginning of Experiment II).

4.2. Cross-Checking: ME Logistic Regression

The complete Time 1 and Time 2 data for Blocks 0 and 1 were submitted to GLMER, with the following model specification.

(10) Response~Time * Type + (1 + Time + Type | Participant) + (1 + Time | ItemCode) + (1 | Version)

Unfortunately, GLMER failed to converge on a fit (even when the RE component for Version had been removed, due to a theta parameter of zero). Given that the primary point of interest concerned the *Whether* items in relation to the grammatically well-formed (Good) items (i.e., because *Whether* was the only sentence type for which the Wilcoxon tests had indicated a significant effect), the dataset was next trimmed to include only the *Whether* and Good sentence types. At that point, using the model specification in (10), GLMER succeeded. The resulting parameters for the fixed effects are shown in Table 2. The results are fully consistent with those from the Wilcoxon tests: the effect of "changing" from the control items to the *Whether* items was a large reduction in acceptance at Time 1 but a significantly smaller reduction at Time 2.

Table 2. Table of fixed effects for Experiment II.

Predictor	Estimate	SE	*Z*	*p*
(Intercept)	3.53	0.82	4.31	<.001
TimeT2	2.21	1.69	1.31	(>.10)
TypeWhether	−7.11	2.76	−2.57	<.05
TypeWhether:TimeT2	7.90	4.01	1.97	<.05

4.3. Follow-Up Testing

As indicated in Section 4.1, when we compare the initial responses (i.e., during the first two blocks of the stimuli) for Time 1 and Time 2, nine of the 15 participants showed an increase in their acceptance of *whether* items, and none showed a decrease. The results were very similar for each version. Of the eight participants who saw Version A at Time 1 and Version B at Time 2, five (i.e., about half) showed an increase from Time 1 to Time 2, and none showed a decrease. Of the six who saw Version B at Time 1 and A at Time 2, three

increased, and none decreased. The one participant who saw Version B at both Time 1 and Time 2 also showed an increase.

In sum, Experiment II indicates that the characteristic of Within-speaker Persistence (3d), reported anecdotally by linguists, also holds (at least in the case of *Whether*-Island violations) for experimentally induced satiation in non-linguists. The increase in "yes" responses on *Whether* items that was statistically significant by the end of Time 1 testing, was still statistically significant four weeks later. Indeed, when we examined performance of the 15 participants individually (Section 4.1), we found that 10 had changed at Time 1 from zero "yes" responses in Blocks 0 and 1 to at least one "yes" response in Blocks 3 and 4 (i.e., they had satiated). In Blocks 0 and 1 of the Time 2 testing, nine of these 10 satiators still said "yes" to at least one *Whether* item. Indeed, given the extremely low likelihood of encountering any *Whether*-Island violations between testing sessions, and given that participants had judged only six examples at Time 1 (five with *wonder whether*, plus a post-test item with *ask whether*), the persistence of the satiation effect is remarkable. This degree of persistence suggests that the satiation the participants had experienced on *whether* items was a "learning" effect rather than a short-term priming effect.

5. Experiment III: Variation in Effect Size

5.1. Overview

A possible concern about Experiments I and II is that they are based on a sample of only 15–20 individuals. This is an especially important consideration given that satiation on CNPCs was weak in (Snyder 2000) (i.e., detected in only five of 20 participants) and not detected at all in Experiments I and II. Increasing the sample size will potentially allow us to reproduce, and better characterize, whatever satiation effect is present for CNPCs.

In Experiment III, the sample size was increased to 151 individuals. The participants were undergraduates taking a large introductory course on the philosophy of language. (None of them had participated in Experiment I or II.) The stimuli were nearly identical to those in Experiments I and II, but they were presented online (one item at a time, so as to control the order in which the judgments were made), and the two items shown in (11) were added to the post-test.

(11) a. *Wh*-Island violation with *wonder why*
(Context: Olga wonders why Sally likes Fred.)
Test Sentence: "Who does Olga wonder why Sally likes?"

b. *Wh*-Island violation with *know how*
(Context: Sue knows how Bill fixed the motorcycle.)
Test Sentence: "What does Sue know how Bill fixed?"

Two fully grammatical items were also added to the post-test. Hence, a participant judged 84 items in total.

Participants began by answering questions about their language background and then were randomly assigned to Version A or B. The initial sample included 194 individuals, but the data were discarded from 29 participants who reported (in answer to the initial questions) that English was not the first language they had acquired. Data were also discarded if a participant's rate of "Yes" responses to fully grammatical items was more than two standard deviations below the group's average or if the rate of "Yes" responses to "deviant" items was more than two standard deviations above the average. Fourteen additional individuals were excluded by these criteria for a final sample of 151.

5.2. Primary Analysis: Wilcoxon Tests

Wilcoxon Signed-Rank Tests indicated possible satiation on four sentence types, namely the sentences violating *Whether*-Island, CNPC, *That*-trace, and Subject-Island constraints. On *Whether* Islands, 70 of 151 participants showed a change between the initial two and the final two blocks, and for 56, it was an increase ($W = -1{,}645$, $n_{s/r} = 70$, $Z = -4.81$, $p < .0001$). For CNPC violations, 36 showed a change, and for 28, it was an increase ($W = -394$, $n_{s/r} = 36$, $Z = -3.09$, $p < .005$). For *That*-trace violations, 72 showed a change,

and for 49, it was an increase ($W = -940$, $n_{s/r} = 72$, $Z = -2.64$, $p < .01$), and for Subject Islands, 60 showed a change, and for 40, it was an increase ($W = -650$, $n_{s/r} = 60$, $Z = -2.39$, $p < .05$). None of the other sentence types showed a significant change.

5.3. Cross-Checking: ME Logistic Regression

Findings were cross-checked using ME logistic regression with the same model structure (9) that was tried initially on the data from Experiment I (i.e., with random intercepts for Version and without any re-scaling of the Block number). The software converged on a model fit, as shown in Table 3, and indicated possible satiation on extraction from *Whether* Islands, Complex NPs, Subject Islands, and *That-trace* environments. Hence, the results were fully consistent with the results from the Wilcoxon tests.

Table 3. Table of fixed effects for Experiment III.

Predictor	Estimate	SE	*Z*	*p*
(Intercept)	3.10	0.22	13.89	<.001
Block	−0.01	0.04	−0.13	(>.10)
TypeAdjunct	−5.28	0.54	−9.75	<.001
TypeCNPC	−7.71	0.70	−10.95	<.001
TypeLBC	−10.72	1.58	−6.80	<.001
TypeSubject	−5.14	0.55	−9.43	<.001
TypeThat	−4.73	0.53	−8.84	<.001
TypeWant	−1.78	0.53	−3.37	<.001
TypeWhether	−4.38	0.52	−8.38	<.001
Block:TypeAdjunct	0.11	0.09	1.29	(>.10)
Block:TypeCNPC	0.53	0.14	3.81	<.001
Block:TypeLBC	−0.28	0.24	−1.19	(>.10)
Block:TypeSubject	0.24	0.10	2.56	<.05
Block:TypeThat	0.23	0.08	2.79	<.01
Block:TypeWant	0.12	0.09	1.45	(>.10)
Block:TypeWhether	0.40	0.08	4.85	<.001

5.4. Follow-Up Testing

The next step was to check whether these cases met the additional criterion discussed above: Did acceptance increase significantly in both versions? For *wonder whether*, findings were fully consistent between versions. Recall that with the two versions combined, 70 participants showed a change in acceptance, and for 56, it was an increase. In Version A, 34 showed a change, with 28 increasing (Wilcoxon Signed-Rank Test, $W = -399$, $n_{s/r} = 34$, $Z = -3.41$, $p < .001$), and in Version B, 36 showed a change, with 28 increasing ($W = -434$, $n_{s/r} = 36$, $Z = -3.41$, $p < .001$). This qualifies as reliable evidence of a satiation effect on *Whether* Islands.

For CNPC violations as well, the findings were fully consistent across versions. Recall that, with the two versions combined, 36 participants showed a change, with 28 increasing. In Version A, 18 showed a change, with 15 increasing ($W = -114$, $n_{s/r} = 18$, $Z = -2.47$, $p < .05$), and in B, 18 showed a change, with 13 increasing ($W = -91$, $n_{s/r} = 18$, $Z = -1.97$, $p < .05$). Note that, in their initial acceptance rate, the CNPC items were quite similar to LBC items. Figure 7 provides a side-by-side comparison of LBC, where Block 0 acceptance was approximately 5% and no satiation was evident, versus CNPC, where Block 0 acceptance was just under 5% and satiation clearly occurred.

In the case of *That*-trace, recall that 72 participants showed a change, and for 49, it was an increase. Yet, this increase was overwhelmingly driven by Version B, where 39 showed a change, and for 30 (i.e., 77%), it was an increase ($W = -444$, $n_{s/r} = 39$, $Z = -3.09$, $p < .005$). On A, however, where 33 showed a change, this was an increase in only 19 (58%) of the cases ($W = -47$, $n_{s/r} = 33$, $Z = -0.42$, $p > .10$ NS). The lack of a significant change in Version A means the findings do not qualify as reliable evidence of satiation. Instead, they were quite possibly an artifact of the particular order in Version B. (For a side-by-side comparison

of *That*-trace with a Block 0 acceptance of approximately 25% and *Whether*-Island violations with a Block 0 acceptance just above 30%, see Figure 8.)

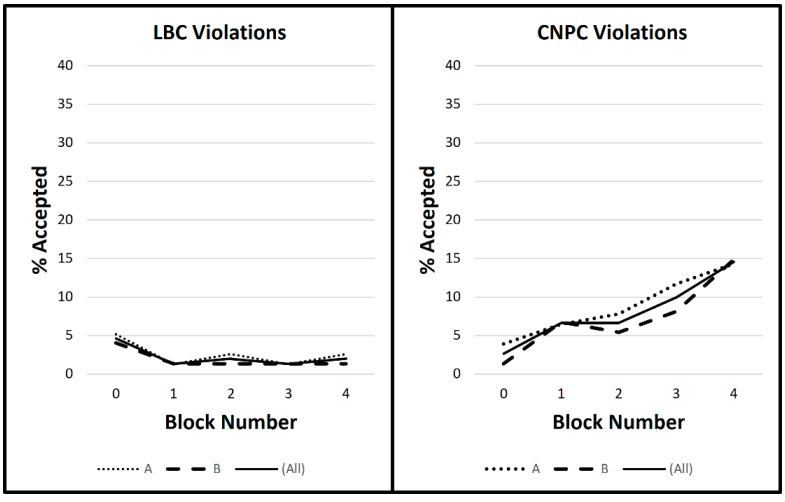


Figure 7. Comparison of LBC violations and CNPC violations in Experiment III.

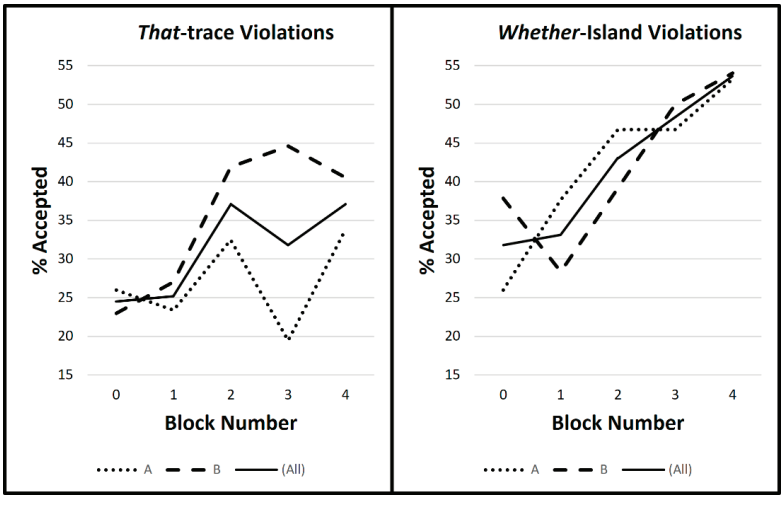


Figure 8. Comparison of *That*-trace and *Whether*-Island violations in Experiment III.

Figures 9 and 10 show the findings for the remaining sentence types. In the case of Subject Islands, recall that 60 participants showed a change, and for 40, it was an increase. When the versions are viewed separately, there is still a significant change in version A ($W = -213$, $n_{s/r} = 30$, $Z = -2.19$, $p < .05$), but the change observed in version B does not reach significance ($W = -116$, $n_{s/r} = 30$, $Z = -1.19$, $p > .10$ NS). The absence of a significant change in Version B means the findings from Experiment III do not qualify as reliable evidence of satiation, but this could well change in a follow-up study (as will be discussed momentarily).

In sum, Experiment III provides clear evidence of satiation on *Whether* Islands and complex NPs but not on the other sentence types examined. The results are entirely

consistent with (Snyder 2000) and largely consistent (i.e., in all respects, except for CNPC) with Experiments I and II. Once again, there was no possibility of response equalization (in the sense of the REH), and yet, a familiar pattern emerged: clear-cut satiation on *Whether* Islands (and, this time, also on CNPC violations, as in Snyder 2000) but not on LBC violations and not on *That*-trace violations. There was also no reliable evidence of satiation on Subject Islands, Adjunct Islands, LBC violations, or *Want-for* environments. Naturally, this does not preclude the possibility that one or more of these latter sentence types will show clear evidence of satiation in another study, especially if the experimental conditions are different. Indeed, in the literature review in Section 7, we will see that satiation is sometimes found for Subject-Island violations but chiefly in studies where participants judged a greater number of examples than they did in Experiments I–III.

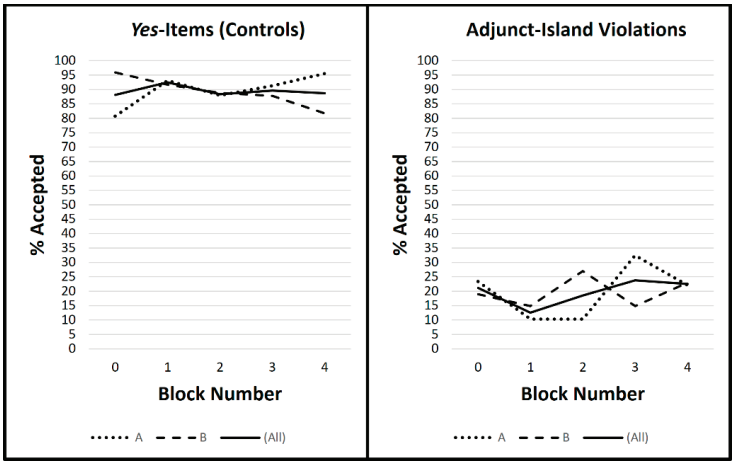


Figure 9. Experiment III, *Yes*-items and Adjunct-Island violations.

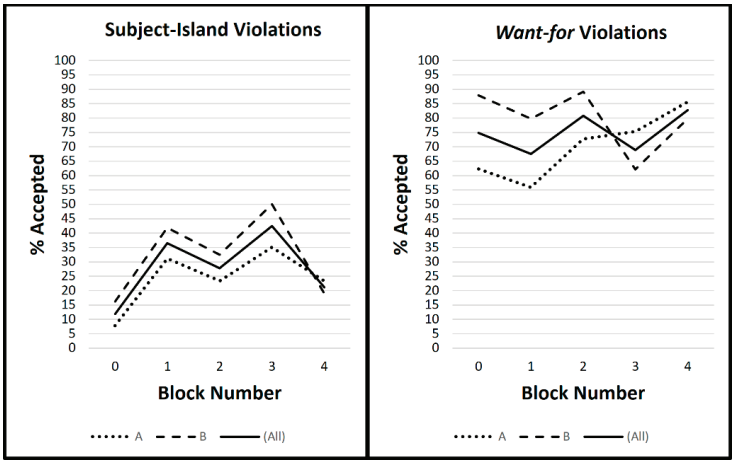


Figure 10. Experiment III, Subject-Island and *Want-for* violations.

5.5. Assessing Effect Size

Finally, a key difference from Experiment I is the clear satiation on CNPC items. The findings of Experiment III actually serve to clarify why this difference exists: in absolute terms, the effect size for CNPC violations was extremely small. With our sample size of

more than 150 participants, we can characterize the effect fairly precisely, and it turns out to have been unrealistic to expect reliable detection of satiation on CNPC items with only 20 participants, as in Experiment I.

In Experiment III, the average acceptance rate for CNPC violations increased from 5% of participants in Blocks 0 and 1 to 12% in Blocks 3 and 4. For present purposes, let's make the (generous) assumption that these rates are a good estimate for the larger population of English-speaking college students. In that case, a simple probability calculation indicates the following. Any participant drawn from this general population and presented with the same materials should have a .013 probability of accepting exactly two more CNPC items in Blocks 3 and 4 than in Blocks 0 and 1. This is because p("No" in Block 0) $\times$ p("No" in Block 1) $\times$ p("Yes" in Block 3) $\times$ p("Yes" in Block 4) = (1 − .05) (1 − .05) (.12) (.12) = .013. Likewise, there should be a .192 probability of accepting one more, a .074 probability of accepting one fewer, and a .00194 probability of accepting two fewer. By power analysis, it follows that N = 76 is the absolute smallest sample size for which the expected frequencies of participants in these four categories (i.e., an expected (76) (.013) = one participant who increases by two, 15 who increase by one, six who decrease by one, and zero who decrease by two) will result in a significant change by Wilcoxon test (at $p < .05$).

The moral is that, even within the set of sentence types that are susceptible to satiation, the strength of the effect may vary as a function of the specific linguistic constraint that is violated. Of the 151 participants in Experiment III, there were 125 who rejected at least one of the two *wonder whether* items in Blocks 0 and 1 and therefore had the possibility of showing increased acceptance (satiation) in Blocks 3 and 4. In fact, 56 of the 125 (45%) satiated. In contrast, 149 of the 151 participants rejected at least one of the CNPC items in Blocks 0 and 1, and only 28 (19%) showed satiation. This raises several questions. For a start, we might ask why, even on *whether* items (where we saw quite a strong satiation effect), fewer than half the participants showed any detectable change. Is this purely a matter of chance, or does susceptibility to the effect relate systematically to some other aspect of an individual's cognitive profile? This would be an interesting question for future research.

At the same time, we might ask whether this dimension of the satiation phenomenon increases its value as a diagnostic tool: perhaps both the susceptibility of a sentence type to satiation in the first place and the strength of the satiation effect observed can provide useful information about the source of the initial unacceptability. This will be taken up again in Section 8.

6. Carryover Effects of Satiation (Experiments I and III)

Recall that (Snyder 2000) found evidence of carryover effects. Restricting attention to those participants who rejected both *wonder whether* items in the first two blocks, the ones who satiated on *wonder whether* (i.e., who accepted at least one of the two exemplars in the final two blocks) were significantly more likely than the others to accept a post-test item involving *ask whether*. Likewise, among those participants who rejected both CNPC items (with *believe the claim*) in the initial two blocks, the ones who accepted at least one CNPC item in the final two blocks were significantly more likely than the others to accept a post-test item involving *accept the idea*. This section presents the corresponding findings from Experiments I and III.[12]

In Experiment I, the participants showed clear satiation on *Whether* Islands (although not on CNPC items), and they judged the same post-test items used in (Snyder 2000). Following the procedure of (Snyder 2000) (described in Section 2), we will restrict our attention to the 15 participants who had rejected both of the *wonder-whether* items in Blocks 0 and 1. Of these 15, three also rejected the *wonder-whether* items in Blocks 3 and 4. Among these three non-satiators, none accepted the post-test item with *ask whether.* In contrast, of the 12 satiators (all of whom accepted at least one of the two *wonder whether* items in Blocks 3 and 4), four (=33%) accepted the post-test item. Hence, the data are consistent with the presence of carryover to the post-test item, although the small numbers make it

difficult to assess the finding statistically. (In particular, the base rate of zero acceptance in the non-satiators makes the use of a binomial test inappropriate.)

When we turn to the data from the larger sample in Experiment III, we find evidence of satiation carryover with both CNPC items and *Whether* Islands. For CNPCs, 139 participants rejected both items in Blocks 0 and 1. Of these 139, 112 were non-satiators (i.e., they also rejected the CNPC items in Blocks 3 and 4), and of these 112, 28 (=25%) accepted the post-test item. In contrast, 15 (=56%) of the 27 satiators accepted it. Hence, there was significant carryover (binomial $p < .005$).

As noted in Section 5, the post-test in Experiment III included an example of *wh*-extraction across *ask whether* (as was also the case in Experiment I), plus one example each for *wonder why* and *know how*. Possible carryover from satiation on *wonder whether* was checked for all three of these post-test items. For *wonder whether*, 79 participants rejected both examples in Blocks 0 and 1. Of these 79, 38 were non-satiators: three (8%) accepted *wonder why*, eight (21%) accepted *know how*, and six (16%) accepted *ask whether*. Of the 41 satiators, 11 (27%) accepted *wonder why*, 10 (24%) accepted *know how*, and nine (22%) accepted *ask whether*. Application of the binomial tests yields $p < .001$ for *wonder why* but $p > .10$ for both *know how* and *ask whether*. The lack of a significant carryover effect for *ask whether* is a departure from (Snyder 2000).

To sum up, in Experiment III (as in Snyder 2000), there was significant carryover from CNPC items involving *believe the claim* to items with *accept the idea*. Yet, the findings for *wh*-islands were more complex. On the one hand, there was significant carryover from *wonder whether* to *wonder why*, which is interesting insofar as it suggests the satiation on *Whether* Islands may be independent of the many special characteristics of the English *wh*-complementizer *whether*. Yet, in Experiment III, there was no significant carryover from *wonder whether* to *ask whether* (as there had been in Snyder 2000 and possibly also in Experiment I), nor was there significant carryover to *know how*. Hence, there is clearly a need for additional research.

Before concluding this section, one final point should be examined. Given that Experiment III yielded evidence of satiation on both *wonder-whether* and CNPC items, we can ask about the relationship between the two within individual participants. Did participants who satiated on one necessarily also satiate on the other? In particular, did individuals in the smaller set of participants who satiated on CNPCs necessarily also accept *wonder-whether* items by the end of the experiment?

The answer is "no". Overall, there were 55 individuals who satiated on *wonder whether* and 29 who satiated on CNPCs. There were only 14 individuals in the intersection. In other words, there were 15 individuals who satiated on CNPCs but showed no increase in their acceptance of *wonder whether*, and there were 41 individuals who satiated on *wonder whether* but showed no increase in their acceptance of CNPCs. Moreover, five of the individuals who satiated on CNPCs actually rejected *wonder whether* entirely (i.e., in both of Blocks 3 and 4). Hence, the satiation effects for these two sentence types appear to be independent.

7. Comparison of Findings across Studies

The work in (Snyder 2000) has given rise to a substantial, growing literature on satiation, which includes new findings from experimental studies (e.g., Hiramatsu 2000; Braze 2002; Francom 2009; Goodall 2011; Crawford 2012; Maia 2013; Christensen et al. 2013; Hofmeister et al. 2013; Chaves and Dery 2014; Do and Kaiser 2017), as well as some efforts to apply these findings to issues in theoretical syntax (e.g., Boeckx 2003, Chapter 3; Stepanov 2007).[13] The present section reviews this literature to assess the consistency of findings across studies. Attention will be focused on studies examining one or more of the same English sentence types studied in (Snyder 2000) (and, hence, Experiments I–III), in order to identify the possible effects of methodological differences across studies.[14] (For other recent surveys of the satiation literature, see Sprouse and Villalta 2021; Snyder 2021.)

7.1. Response Equalization?

First, while Experiment I was the most direct test to date of Sprouse's (2009) REH critique of (Snyder 2000) (as discussed in Section 2, note 7), the conclusions were certainly anticipated by others in the literature. For example, the following four researchers had all reported satiation effects in experiments with a perfect balance of "expected YES" and "expected NO" items: Hiramatsu (2000, Experiment II - henceforth "E2"), Braze (2002), Francom (2009, E2), and Crawford (2012).[15]

7.2. Consistency of Findings and Points of Variation

For the English sentence types examined in Experiments I–III, the results are very much in-line with other studies in the literature. (A synopsis is provided in Tables 4 and 5.) A clear majority detect satiation on argument extraction from *Whether* Islands.[16,17] Sprouse's (2009) A3 and B1 and 2 are exceptions and will be discussed in Section 7.3.1.[18]

Table 4. Sentence types on which satiation has been induced experimentally. ("ME" = magnitude estimation; "Y/N" = yes/no task.).

Sentence Type	Satiation?	Experiment	Context Sentence?	N	Exposures	Task
Whether Island	Yes	(Braze 2002)	Yes	35	9	Y/N
	Yes	(Crawford 2012)	Yes	22	7	Scale
	Yes	(Francom 2009: E1)	No	205	5	Y/N
	Yes	(Hiramatsu 2000: E1)	Yes	33	7	Y/N
	Yes	(Hiramatsu 2000: E2)	Yes	11	7	Y/N
	Yes	(Snyder 2000)	Yes	22	5	Y/N
	Yes	*(Experiment I)*	Yes	20	5	Y/N
	Yes	*(Experiment III)*	Yes	151	5	Y/N
	No	(Sprouse 2009: A3)	No	20	10	ME
	No	(Sprouse 2009: B1)	No	25	5	Y/N
	No	(Sprouse 2009: B2)	No	19	5	Y/N
CNPC	Yes	(Goodall 2011)	Yes	45	5	Y/N
	Yes	(Snyder 2000)	Yes	22	5	Y/N
	Yes	*(Experiment III)*	Yes	151	5	Y/N
	No	(Hiramatsu 2000: E1)	Yes	33	7	Y/N
	No	(Sprouse 2009: A4)	No	17	10	ME
	No	(Sprouse 2009: A5)	Yes	20	10	ME
	No	(Sprouse 2009: B1)	No	25	5	Y/N
	No	(Sprouse 2009: B2)	No	19	5	Y/N
	No	*(Experiment I)*	Yes	20	5	Y/N
Subject Island	Yes	(Chaves and Dery 2014: E1)	No	60	20	Scale
	Yes	(Chaves and Dery 2014: E2)	No	55	14	Scale
	Yes	(Francom 2009: E1)	No	205	5	Y/N
	Yes	(Francom 2009: E2)	No	22	8	Y/N
	Yes	(Hiramatsu 2000: E1)	Yes	33	7	Y/N
	Yes	(Hiramatsu 2000: E2)	Yes	11	7	Y/N
	No	(Crawford 2012)	Yes	22	7	Scale
	No	(Goodall 2011)	Yes	45	5	Y/N
	No	(Snyder 2000)	Yes	22	5	Y/N
	No	(Sprouse 2009: A1)	No	20	14	ME
	No	*(Experiment I)*	Yes	20	5	Y/N
	No	*(Experiment III)*	Yes	151	5	Y/N

Table 5. Sentence types on which studies have consistently failed to induce satiation. ("ME" = magnitude estimation; "Y/N" = yes/no task.).

Sentence Type	Satiation?	Experiment	Context Sentence?	N	Exposures	Task
Adjunct Island	No	(Braze 2002)	Yes	16	9	Y/N
	No	(Crawford 2012)	Yes	22	7	Scale
	No	(Francom 2009: E1)	No	205	5	Y/N
	No	(Francom 2009: E2)	No	22	8	Y/N
	No	(Goodall 2011)	Yes	45	5	Y/N
	No	(Hiramatsu 2000: E1)	Yes	33	7	Y/N
	No	(Hiramatsu 2000: E2)	Yes	11	7	Y/N
	No	(Snyder 2000)	Yes	22	5	Y/N
	No	(Sprouse 2009: A2)	No	24	14	ME
	No	(Sprouse 2009: B1)	No	25	5	Y/N
	No	(Sprouse 2009: B2)	No	19	5	Y/N
	No	*(Experiment I)*	Yes	20	5	Y/N
	No	*(Experiment III)*	Yes	151	5	Y/N
Left Branch	No	(Francom 2009: E1)	No	205	5	Y/N
	No	(Francom 2009: E2)	No	22	8	Y/N
	No	(Goodall 2011)	Yes	45	5	Y/N
	No	(Hiramatsu 2000: E1)	Yes	33	7	Y/N
	No	(Hiramatsu 2000: E2)	Yes	11	7	Y/N
	No	(Snyder 2000)	Yes	22	5	Y/N
	No	(Sprouse 2009: B1)	No	25	5	Y/N
	No	*(Experiment I)*	Yes	20	5	Y/N
	No	*(Experiment III)*	Yes	151	5	Y/N
That-trace	*(See text)*	(Hiramatsu 2000: E1)	Yes	33	7	Y/N
	No	(Francom 2009: E1)	No	205	5	Y/N
	No	(Francom 2009: E2)	No	22	8	Y/N
	No	(Goodall 2011)	Yes	45	5	Y/N
	No	(Snyder 2000)	Yes	22	5	Y/N
	No	*(Experiment I)*	Yes	20	5	Y/N
	No	*(Experiment III)*	Yes	151	5	Y/N
Want-for	*(See text)*	(Hiramatsu 2000: E1)	Yes	33	7	Y/N
	(See text)	(Francom 2009: E1)	No	205	5	Y/N
	No	(Snyder 2000)	Yes	22	5	Y/N
	No	*(Experiment I)*	Yes	20	5	Y/N
	No	*(Experiment III)*	Yes	151	5	Y/N

Two other sentence types have sometimes, but not always, shown satiation: CNPCs and Subject Islands. Studies testing CNPCs include Goodall 2011, which found clear satiation, as well as several others that did not.[19] As discussed in Section 5, Experiment III sheds considerable light on this variability; it appears the effect size for CNPCs is much smaller than for *Whether* Islands. Without a sizable number of participants (a bare minimum of $N = 76$, it seems, for the specific design and materials used in Experiment III), there is a high probability of failing to detect satiation on CNPCs (i.e., even if some degree of satiation is occurring). As seen in Table 4, two of the three experiments finding significant satiation on CNPCs (including Experiment III above) had at least 40 participants, while those not finding it all had fewer than 40.

Note that (Sprouse 2009) included four experiments trying to induce satiation on CNPCs, each with 25 or fewer participants. Individually, these experiments probably had little chance of succeeding, but overall (with 81 participants in total), the chances of detecting it (at least once) were perhaps not so bad. The larger issue may have been that the stimuli in all but one of these experiments (A5) omitted the context sentence, with the result that there was no clear indication of the intended meaning. This looks like it might have been a critically important change, because, in Table 4, the experiments that succeeded all provided context sentences. If so, the fact that only A5 included context sentences, together with the fact that A5 had only 20 participants, may explain the absence of satiation in Sprouse's experiments.

Turning to Subject Islands, ten of the other studies in Table 4 tested argument extraction from DPs in the subject position (especially DPs that were underlyingly direct objects, as

with passives and unaccusatives). Six found satiation, and four did not. One problem in some of the latter studies may have been an insufficient number of exposures. Almost all the studies finding satiation (five out of six) increased the number of exposures beyond the original five in (Snyder 2000).[20] In fact, Hiramatsu (2000, E1), who employed seven exposures, noted that the satiation evident in Block 7 was not yet detectable in Block 5.[21]

For other sentence types examined in Experiments I–III above, the absence of detectable satiation is also largely consistent with the literature (see Table 5). For the LBC, seven other studies tested for satiation, and none found it. Adjunct Islands were checked in eleven other studies, and again, none found satiation.

That-trace violations have not in general shown satiation, but more needs to be said. Sprouse (2009, p. 331, Table 2) characterized (Hiramatsu 2000) as having found satiation on *That*-trace, but the situation was unclear. Hiramatsu (p. 107) expressed concerns about the quality of her data for *That-trace* and *Want-for* (which were tested only in E1). She reported that multiple participants had eventually begun crossing out the word *that* or *for* and then marking "Yes". Moreover, on p. 111, she seems to disavow the data for these sentences altogether: "As we saw in the previous section, we do not have a clear picture of the results for [...] *That-trace* and *Want-for* sentences." Hence, the cautious approach would be to set those findings aside, and the other studies of *That-trace* in Table 5 found no satiation.

In the case of *want-for*, once we set aside (Hiramatsu 2000), the main data in the literature (aside from Snyder 2000, which found no satiation, and Experiments I–III above, which likewise found no satiation) come from Francom (E1), who does report satiation. As it happens, Francom employed Snyder's (2000) method of counterbalancing the order of presentation. He did not originally provide information about the consistency of responses across the two orders, but he very kindly shared his data. This made it possible to check whether the change in acceptance was comparable across the two versions.

As it turned out, the evidence for satiation on *want-for* did not satisfy this criterion. Collapsing across the two versions, acceptance increased from 75% in Blocks 1 and 2 to 83% in Blocks 4 and 5 ($W = -856$, $n_{s/r} = 61$, $Z = -3.08$, $p < .005$). Yet, this change was driven almost entirely by participants receiving Version B.[22] The acceptance on Version B shifted from 71% to 83% ($W = -390$, $n_{s/r} = 36$, $Z = -3.06$, $p < .005$), but the acceptance on Version A went only from 79% to 83% ($W = -86$, $n_{s/r} = 25$, $Z = -1.15$, $p > .10$ NS). Hence, the increased acceptance of *want for* at the end of Francom's experiment was probably due to an accidental property of the presentation order in Version B. By the criteria used in Experiments I–III (specifically, the requirement for the effect to be present in both orders of presentation), the findings do not qualify as reliable evidence of satiation.

In sum, across the studies reviewed here, the sentence types showing a satiation effect have consistently been some combination of *Whether*-Island, CNPC, and Subject-Island violations. At least by the criteria employed here, no study has yielded reliable evidence of satiation on Adjunct-Island, Left-Branch, *That*-trace, or *Want-for* violations.

7.3. Points of Variation in Method

7.3.1. Experimental Set-Up

Studies attempting to induce satiation have varied somewhat in their experimental set-up. As can be seen in Tables 4 and 5, one potentially important variable is whether a context sentence was provided. The studies that included a context sentence have mostly succeeded in inducing satiation, at least for *Whether* Islands, but the results have been less consistent when it was omitted.[23]

Note that providing a context sentence is one way of conveying the intended meaning of a sentence. Arguably, judgments of linguistic acceptability are always (at least implicitly) relative to an interpretation. For example, the acceptability of the English sentence *John likes him* depends critically on whether *him* is taken to mean the same person as *John*; hence, referential indexing is provided in the literature on binding theory. In other cases, one does find linguists simply placing an asterisk on a sentence without specifying an intended meaning, but in practice, this appears to mean one of two things. Either the sentence is

unacceptable on what the linguist takes to be the "obvious" interpretation or the linguist believes the sentence is unacceptable no matter what the intended meaning is. Thus, in an experimental study of linguistic acceptability, one possible effect of including a context sentence is simply facilitation of the judgment task by making it easier for the participant to identify an intended meaning when making the judgment.

Yet, as suggested by an anonymous reviewer, the inclusion of a context sentence (and thus, clarification of the intended meaning) might have an additional, quite important role that would be specific to an experiment on satiation effects. This is because it helps the participant identify one particular way of parsing the test sentence. As will be discussed in Section 8, there could be a number of relevant consequences. For one, having this information could lead a probabilistic parser to increase the expected probability of an uncommon parse (e.g., in the context of *wh*-extraction from a subject island, the probability of a parse positing a gap inside the subject of an embedded clause). Another effect could be helping the participant recognize that adopting a "marked" syntactic option will render the sentence grammatically possible (e.g., in the context of *wh*-extraction from a CP inside an NP, adopting the option—which is potentially a marked option—of treating the CP as a complement to N rather than simply an N-bar adjunct). Thus, there are good reasons to expect that the inclusion of a context sentence might facilitate satiation and, moreover, that the facilitation might apply to certain sentence types more than others.

Another salient point of variation across different satiation studies is the nature of the judgment task: Does the participant provide a Yes/No judgment, a rating on a numerical scale, or an estimate of magnitude? Most studies that successfully induced satiation employed a Yes/No task, although Crawford 2012; Chaves and Dery 2014 (E1–2) employed a numerical scale. Sprouse 2009 (A1–5) differed in choosing magnitude estimation (ME). At present, it is unclear whether the choice of task affects the findings for satiation—and, if so, why this would be the case. (For a recent discussion of the task characteristics of ME, see Featherston 2021 and the references therein.) What is needed is a side-by-side comparison of these methods within a single satiation study.

As already noted, two other variables appear to be critically important: the number of exposures to each sentence type and the number of participants in the study. The information in Table 4 suggests that satiation on Subject Islands is difficult to obtain, unless the number of exposures is at least seven, and that satiation on CNPCs is likewise difficult to obtain, unless the number of participants is substantial (at bare minimum 76 for the specific materials and design in Experiment III). These points will be taken up again in Section 8.[24]

7.3.2. Variation in the Stimuli

Another salient point of variation concerns the detailed syntax of the test sentences. For example, Hiramatsu 2000 contrasted two types of Subject-Island violations, involving extraction from a subject DP that was either the underlying object or the underlying subject of a transitive verb. Interestingly, she found satiation only with underlying objects. In a similar vein, she contrasted the extraction of arguments versus adjuncts from a *Whether* Island and found satiation only for arguments.

7.3.3. Variation in Data Analysis

Studies have varied in their statistical methods, but the differences seem to be immaterial. Francom (2009, pp. 32–35) applied sign tests, paired *t*-tests, ANOVA, and logistic regression, with identical results. Likewise, the datasets from Experiments I–III in this paper were analyzed both with ME logistic regression and with a more traditional method (the Wilcoxon Signed-Rank Test), and the results were effectively identical.

In contrast, what is clearly of great importance is confirming that one's data are internally consistent: Do we see the consistency across orders of presentation that we should for an effect at the level of grammatical structure? As illustrated in Experiment III, this common-sense check can have a critically important influence on the conclusions

drawn. Furthermore, in Section 7.2, they helped eliminate an apparent conflict across studies in the findings for *want-for*.

8. Directions for Future Research

8.1. Satiation as a Diagnostic Test

We now return to the question of how investigating satiation could benefit generative linguistics. First, a number of potentially valuable ways to apply our current knowledge of satiation might follow from a proposal made by Goodall (2011, p. 35):

> [I]f one unacceptable sentence type is satiation-inducing and another is not, it is unlikely that their unacceptability is attributable to the same underlying principle. This suggests, for instance, that violations of *whether* islands, which are susceptible to satiation, and *that*–trace violations, which are not, must be due to different underlying principles, in accord with the general consensus in the literature about these two phenomena.

Following this line of reasoning, and incorporating the findings discussed in this article, one can see a number of immediate applications. Whenever a linguistic theory (be it a theory in syntax, semantics, or morphophonology) posits a single source for the unacceptability of two different sentence types X and Y, testable predictions immediately follow.

For example, one possibility will be to run a pair of studies modeled on Experiment III. In one of the studies, we add a single example of sentence type X to each block. In the second study, we use examples of Y in place of X. Upon completion, we check to see if X and Y are alike (or disparate) in whether their initial unacceptability satiates. If satiation is present for both, we can also check whether the number of exposures required for satiation is comparable across X and Y, and we can check whether the percentage of participants who show a change in their judgment is comparable across X and Y.

If the satiation findings for X and Y are either highly similar, or highly dissimilar, the interpretation will be straightforward. More complex (and, no doubt, more interesting) will be the intermediate cases, where some of the diagnostics come out as expected under the hypothesis of a single source of unacceptability and others do not. This sort of mixed case might, for example, indicate that X and Y overlap only partially in the factors rendering them (initially) unacceptable.

Yet, there are a number of ways for the ideas just sketched to be too simplistic. In particular, there is a tacit assumption that the underlying source of unacceptability is the thing undergoing change. Suppose, for example, that a specific UG constraint on syntactic movement is what is rendering both X and Y unacceptable. If this constraint is somehow weakened by satiation, then both X and Y should become more acceptable. However, suppose that the UG constraint is immune to satiation and that something else is changing. For example, perhaps a speaker can learn to reanalyze structure X as a superficially similar but syntactically distinct structure X-prime, to which the UG constraint does not apply. If the reanalysis operation depends on surface characteristics that are present on X but absent from Y, only X will be able to satiate, even though the cause of the initial unacceptability of X and Y is exactly the same.

8.2. Explaining Satiation

Before we try to use satiability as a diagnostic, we will naturally want to know as much as we can about what exactly satiation is. A logical starting point is to ask whether satiation is a unitary phenomenon. Is there essentially the same process at work in every example of a sentence type that satiates (according to the operational definition of satiation in 3)? Alternatively, are there different mechanisms at work in different sentence types?

The findings in this article can at least help narrow down the possible answers. Consider the following "strongly unitary" scenario:

> ***Scenario 1.*** Suppose that a kind of "mental alarm" goes off whenever a person's language-processing mechanisms are forced to postulate a grammatically deviant structure for a linguistic expression. Let's assume that the alarm system is highly

> similar from one speaker to another; the strength of the alarm varies along a single, smoothly continuous dimension, and violations of different grammatical constraints all trigger the same alarm, although the strength of the resulting alarm signal may vary with the type of violation. If so, satiation could perhaps be a kind of habituation effect: perhaps repeatedly experiencing a certain level of alarm, over a certain period of time, can make one tolerant.

Under Scenario 1, no matter which sentence types undergo satiation, the mechanism is exactly the same: habituation to alarm signals of a certain magnitude. Grammatical constraints associated with a weak signal should always satiate prior to constraints with a stronger signal. Indeed, satiation on a constraint with a strong signal should yield satiation not only on sentences violating that particular constraint but also on sentence types yielding weaker signals, even if those sentence types violate different constraints and even if those sentence types have never actually been encountered.

The evidence presented in this article speaks against an account along these lines. Specifically, the fact that satiation on CNPC violations in Experiment III was found in a much smaller percentage of participants than satiation on *wonder whether* more or less forces us to conclude, under Scenario 1, that CNPC violations elicit a "louder" alarm signal than *wonder whether* violations. Hence, there is a strong prediction that every single individual who satiated on CNPC violations by the end of Experiment III must have ended up accepting *wonder whether* items as well. At the end of Section 6, however, it was noted that five of the 29 participants who satiated on CNPC violations actually rejected both of the *wonder whether* items in Blocks 3 and 4.

In place of a strongly unitary account, one might consider a "weakly unitary" account along the following lines:

> ***Scenario 2.*** Suppose the language processor has a number of distinct alarm signals, each of which indicates the violation of a different grammatical constraint. In this case we might once again imagine that satiation results from habituating to an alarm signal (and, hence, that satiation is unitary in a certain sense), but now, satiation will proceed independently for different grammatical constraints (i.e., as a separate process of habituation for each of several different alarms). Satiation on a given constraint will require exposure to sentences violating that particular constraint.

Note that, under Scenario 2, the number of exposures required before full habituation occurs might still vary as a function of the constraint in question if (for example) some constraints have "louder" alarms than others.

Is Scenario 2 compatible with the evidence from Experiment III? This depends on our assumptions. If we assume—as seems fair—that, prior to the experiment, the participants had no exposure to either CNPC violations or *wonder whether* violations, and if we assume that each exposure during the experiment is equally effective at promoting habituation, then the same prediction that defeated Scenario 1 will probably exist for Scenario 2. Specifically, by the end of the experiment, every participant will have encountered the same number of CNPC violations and *wonder whether* violations; if that number (of CNPC violations) is sufficient to create habituation on the alarm signal for CNPCs (again assuming that these are the more difficult sentence type to satiate), then the same number (of *wonder whether* violations) should be sufficient to produce habituation on the (distinct, but weaker) alarm signal for *wonder whether*.

Yet, the prediction will change if we assume, for example, that habituation to a given alarm signal requires not only some number of encounters with relevant examples but also some particular internal state in the experimental participant (perhaps something like introspective awareness) that fluctuates from moment to moment. In this case, it might be possible, simply by chance, for a participant to have "genuinely" experienced a smaller number of alarm signals for *wonder whether* violations by the end of the experiment than alarm signals for CNPC violations.

In any case, a complete alternative to Scenarios 1 and 2 would be Scenario 3, which sketches a strongly non-unitary model.

> ***Scenario 3.*** Suppose that different satiable constraints may owe their satiability to different mechanisms. Perhaps, in some cases, satiation results from habituation to a particular constraint's alarm signal, but in other cases, it results from, say, discovering an alternative syntactic analysis of a particular sentence type. For example, perhaps CNPC violations involving *wh*-extraction across *...believe the claim that...* are usually assigned an "unmarked" structure in which the CP is treated as an appositive (i.e., an N-bar adjunct), but UG also permits another, more marked analysis (at least for epistemic nominals, like *claim* and *idea*) in which the CP is a complement selected by N. In terms of Chomsky's (1986) *Barriers* system, the appositive analysis forces the *wh*-phrase to cross two barriers (the lower CP, which is not L-marked, as well as the NP above it, which is a barrier by inheritance). In contrast, no barrier will be crossed if the lower CP is selected by the N. Hence, in this case, satiation is not habituation but, rather, the discovery of a new, UG-compatible (but "marked") parse, which (by hypothesis) was not being exploited before.

Under Scenario 3, it is perhaps less surprising (than under Scenarios 1 and 2) to find participants who have satiated on CNPC violations but who still firmly reject *wonder whether* violations. If satiation on CNPC violations results from a sudden (tacit) insight into UG-compatible structures but satiation on *wonder whether* violations results (say) from the gradual accumulation of a particular volume of experience over time, then an individual can easily satiate on one and not the other.[25]

At this point, it is interesting to note that Chaves and Dery (2018) proposed an explicit model of satiation on Subject Island (SI) violations, and their model seems to be far more compatible with Scenario 3 than Scenarios 1 and 2. This is because their work does not address satiation on sentence types other than SIs, and the proposed mechanism of satiation appears to be specific to SIs. In brief, Chaves & Dery argued that SI violations are not ungrammatical but merely difficult to parse. They assumed the parsing difficulty results from "the fact that subject-embedded gaps are pragmatically unusual—as the informational focus does not usually correspond to a dependent of the subject phrase—and are therefore highly contrary to comprehenders' expectations about the distribution of filler gap dependencies" (Chaves and Dery 2018, p. 1). In their view, comprehenders' expectations can change fairly rapidly with exposure to clear examples of subject-embedded gaps.

Thus, the Chaves–Dery mechanism seems like a plausible candidate for a source of satiation that is specific to SI violations. Let's suppose this proposal is correct for SI violations. Then, as suggested above, perhaps satiation on CNPC violations will turn out to involve discovering a new, UG-compatible (but ordinarily nonpreferred) parse for a CP following an epistemic nominal. Perhaps satiation on extraction from certain *wh*-islands will turn out to involve habituating to a mental "alarm" triggered by a certain type of grammatical violation. This type of non-unitary scenario leads to distinctive predictions, such as the strict absence of satiation carryover effects between sentences of these three types. Experimental tests of such predictions would be a reasonable next step for research into the nature of satiation.

In conclusion, Experiments I–III have provided evidence (i) that experimentally induced satiation, like the satiation that sometimes affects linguists, is restricted to a small, stable set of sentence types; (ii) that, after satiation on one sentence type (e.g., *wh*-movement across *...wonder whether...* or *...believe the claim...*), acceptability sometimes increases for distinct but syntactically related sentence types (such as *...wonder why...* or *...accept the idea...*); (iii) that, for sentence types susceptible to satiation, the difficulty of inducing it (e.g., number of exposures required) varies systematically; and (iv) that, much as satiation in linguists persists over time, experimentally induced satiation (at least in the case of *wonder whether*) can persist for at least four weeks. These findings may suggest an eventual role for satiation in determining whether the perceived unacceptability of two distinct

sentence types has a common source, but more immediately, they suggest that satiation may be a powerful tool for examining the tacit mental operations that are responsible for our judgments of linguistic (un)acceptability.

Funding: This work was supported by the National Science Foundation under NSF IGERT DGE-1144399 and grant DGE-1747486.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the University of Connecticut.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Anonymized data are available from the author.

Acknowledgments: The author is grateful for helpful discussions with (among others) Dave Braze, Rui Chaves, Jean Crawford, Jerrod Francom, Grant Goodall, Kazuko Hiramatsu, and Whit Tabor. The author also thanks the anonymous reviewers for their numerous astute suggestions.

Conflicts of Interest: The author declares no conflict of interest. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Notes

1 The experimental findings published in (Snyder 2000) were first presented as part of (Snyder 1994).

2 The constraints operating in (6a–d,g) were first characterized in (Ross 1967). Early discussions of *That*-trace and *Want-for* effects (6e,f) can be found in (Perlmutter 1968) and (Rosenbaum 1967), respectively.

3 One might be surprised by the the reported lack of satiation on *that*-trace violations, given Sobin's (1987, et seq.) claim that certain varieties of English lack this constraint altogether. Yet, one needs to proceed with caution here, because more recent work has called Sobin's claim into question (see, in particular, Chacón 2015; Cowart and McDaniel 2021).

4 Note that Snyder (2000) measured satiation on a given sentence type by examining (i) the number of participants who accepted more tokens at the end (i.e., final two blocks) of the experiment than at the beginning (initial two blocks) and (ii) the number who accepted more tokens at the beginning than the end. Participants who accepted equal numbers of tokens at the beginning and end were set aside. The assumption was that random variability in judgments is equally likely to create an increase or a decrease. Satiation was detected, for a given sentence type, when there was a statistically significant preponderance (among those whose rate of acceptance changed) of participants who accepted more tokens at the end. Statistical significance was assessed by means of a two-tailed binomial test based on the null hypothesis that increases and decreases each have a 0.5 probability.

5 In (Snyder 2000, p. 579) some of the numbers reported for the carryover effects in *Whether* Islands and CNPC violations were transposed. In the calculations reported here, these errors have been corrected, and the impact is minimal: the two carryover effects that were reported as statistically significant in (Snyder 2000) remain significant after the corrections.

6 These considerations will play a critical role in the approach to data analysis in Experiments I–III below.

7 To foreshadow the findings, the REH will receive no support in Experiment 1. Indeed, as an anonymous reviewer pointed out, there has never been any evidence directly supporting the REH (even in Sprouse 2009). The experiments presented in (Sprouse 2009) had made the change (from Snyder 2000) of perfectly balancing the number of expected yes versus expected no answers, and the participants did not show satiation. The REH was proposed as an explanation for this difference in results, but surprisingly, Sprouse did not report the natural follow-up study of increasing the number of expected "no" answers and showing that the (apparent) satiation of (Snyder 2000) resurfaced. Moreover, as will be discussed in Section 7, Sprouse made additional changes to Snyder'smethodology that could plausibly account for the replication failure. Thus, Experiment 1 is the first direct test of the REH as an explanation for the findings of (Snyder 2000); all aspects of the experiment are exactly the same as in (Snyder 2000) except for a perfect balance of expected "yes"and expected "no" items.

8 Participants, on average, said "yes" to 90% of the grammatical items (standard deviation 8.2). The excluded participants had each answered "yes" to approximately 70%.

9 An anonymous reviewer noted that the acceptance rates for "yes" items, shown in the right panel of Figure 1, are not at the ceiling but, rather, vary somewhat across the different blocks of the experiment. Importantly, this variability will be controlled for statistically in the logistic regression model described in Section 3.6. As noted in Section 3.2, ME logistic regression will be conducted with one level of each factor specified as a baseline for use in "treatment contrasts" (i.e., pairwise comparisons) with each of the other levels of that factor. For Type, the baseline level will be "Good" (i.e., within each block, the results for the seven fully grammatical "yes" items). Hence, whatever rate of acceptance the "yes" items might receive in a given block, that same rate will be the comparison point for each of the other sentence types in that block.

[10] Specifically, values of the theta parameters for the RE structure were obtained by executing getME(Data.model,"theta"), and this revealed a value of zero for Version.(Intercept).

[11] Note: throughout this article, *p*-values are two-tailed.

[12] Experiment II is being set aside here, because the participants had already seen the same "post-test" items earlier in Experiment I.

[13] Much, but not all, of this work focuses on English. Goodall (2011) performed a cross-linguistic comparison of satiation in English versus Spanish (although the findings reviewed here will be limited to his English data). Christensen et al. (2013) detected possible satiation on *wh*-islands in Danish. Maia (2013) discussed a study that he conducted with Wendy Barile finding satiation on *wh*-in-situ (within islands) in Brazilian Portuguese. Interestingly, this final study compared judgments of undergraduates who had recently completed a syntax course covering island effects versus students who had never studied linguistics.

[14] Due to the decision to focus on satiation studies looking at the same sentence types examined in (Snyder 2000), some work on another sentence types, namely superiority violations, will receive only the following remarks. Briefly, Hofmeister et al. (2013) for English and Brown et al. (2021) for both German and English reported increased acceptance of superiority violations after multiple judgments. Yet, there are some reasons to be skeptical that their findings resulted from "syntactic satiation" in the sense intended here—in other words, from the type of satiation reported anecdotally among professional linguists. First, Brown et al. reported that the change in acceptance occurred very rapidly (appearing on the second exposure); second, for both studies, the increase in acceptance was slight, in absolute terms; and third, Brown et al. reported that the same slight increase was found across a range of anomalous sentences that varied in their grammatical structure but that were similar in initially having an intermediate level of acceptability. For the benefit of future investigations, one other point bears mentioning: in both studies, the researchers apparently decided to omit the context sentences of (Snyder 2000). As will be discussed below, the same change may have been responsible for the absence of an expected satiation effect in some of the studies that are reviewed here.

[15] Note that Braze (2002) argued for the existence of a counterpart to satiation in sentence-processing based on an eye-tracker study that he ran in conjunction with an off-line judgment study (with different participants). The findings cited in this section come from the offline study. Yet, if Braze is correct about the sentence-processing counterpart, it both speaks against an account in terms of response equalization (since no judgment of acceptability was elicited) and also has important implications for the project of explaining satiation. (Some related topics will be taken up in Section 8.)

[16] In the case of (Hiramatsu 2000), the discussion here concerns only the results from participants who met her stated inclusionary criterion, namely answering at least 90% of the filler and control items, as expected.

[17] Francom (2009) included a total of five experimental studies, but E1 and 2 seemed to be the most directly comparable with the other studies in this section and, hence, will be the focus.

[18] To disambiguate, Experiments 1–5 from Section 3 of (Sprouse 2009) are prefixed with an "A", and Experiments 1 and 2 from Section 4 are prefixed with a "B".

[19] Francom (2009) (E1 and 2) also reported results for "CNPC violations", but he included a wide range of sentence types under this label, as can be seen from the stimulus lists that he provides in the appendices (pp. 103, 105, 108). Given that the relevance is unclear for "CNPC violations" in the sense intended here, those studies are omitted from the CNPC section of Table 4.

[20] Note that Sprouse, in A1, increased Subject-Island exposures from 5 to 14 but omitted the context sentences. Either this lack of context sentences or his use of a Magnitude Estimation task could (in principle) be responsible for the absence of a satiation effect.

[21] Aside from Sprouse's A1, most studies providing seven or more exposures found satiation on Subject Islands, at least when the subject DPs were underlying objects. The exception is (Crawford 2012), where the 22 participants received seven blocks of exposure (together with context sentences), but no satiation was detected, even for extraction from the subjects of unaccusatives.

[22] To maintain consistency with the earlier discussion of (Snyder 2000) and Experiments I–III, Francom's versions 1 and 2 will be referred to here as A and B, respectively.

[23] One example of success without the use of context sentences is the satiation on Subject Islands in (Chaves and Dery 2014, E1 and 2). Some features of their study that could (in principle) be relevant include the large number of exposures (14–20), the lack of variety in the stimuli (all of the initially unacceptable items had *wh*-extraction from a Subject Island), and the fact that half of the test items employed the D-linked form *which*.

[24] Note that, in principle, there might, or might not, be a simple trade-off between the number of participants in a study and the number of exposures to a given sentence type. If the effect of increasing the number of exposures is a linear increase in the percentage of participants who experience satiation, then perhaps even a small number of exposures will yield a detectable satiation effect if the number of participants is sufficiently large. Alternatively, a given sentence type might turn out to require a minimum number of exposures before any change occurs, no matter how many people participate. The findings from Experiment 3 perhaps favor the latter scenario, since, even with 151 participants, there was no sign of satiation on Subject Islands within the space of five exposures.

[25] As brought to my attention by an anonymous reviewer, there are predictions about the length of time for which satiation will persist that follow directly from different proposed mechanisms. For example, we might reasonably expect indefinite persistence when satiation is due to a "learning" effect (as in my suggestion about satiation on CNPC violations, sketched under Scenario 3). The appropriate predictions could well be different, however, for a mechanism akin to sensory habituation (as sketched under

Scenarios 1 and 2), and they might be different still for the mechanism that Chaves and Dery proposed for Subject Islands (in terms of changing the probability associated with a given parse in a probabilistic parser). In fact, as noted at the end of Section 4.3, the persistence of satiation on *wonder whether* in Experiment II is strongly suggestive of a learning effect rather than the sort of habituation to an alarm signal (i.e., for subjacency violations or the like) suggested in Section 8.2.

References

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [CrossRef] [PubMed]

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–148. [CrossRef]

Boeckx, Cedric. 2003. *Islands and Chains: Resumption as Stranding*. Amsterdam: John Benjamins.

Braze, Forrest David. 2002. Grammaticality, Acceptability, and Sentence Processing: A Psycholinguistic Study. Doctoral dissertation, University of Connecticut, Storrs, CT, USA.

Brown, J. M. M., Gisbert Fanselow, Rebeccah Hall, and Reinhold Kliegl. 2021. Middle ratings rise regardless of grammatical construction: Testing syntactic variability in a repeated exposure paradigm. *PLoS ONE* 16: e0251280. [CrossRef] [PubMed]

Chacón, Dustin A. 2015. Comparative Psychosyntax. Doctoral dissertation, University of Maryland, College Park, MD, USA.

Chaves, Rui P., and Jeruen E. Dery. 2014. Which subject islands will the acceptability of improve with repeated exposure? Paper presented at the 31st West Coast Conference on Formal Linguistics, Tempe, Arizona, USA, 8 February 2013; Edited by Robert E. Santana-LaBarge. Somerville: Cascadilla Proceedings Project, pp. 96–106.

Chaves, Rui P., and Jeruen E. Dery. 2018. Frequency effects in Subject Islands. *Journal of Linguistics* 55: 475–521. [CrossRef]

Chomsky, Noam. 1986. *Barriers*. Cambridge: MIT Press.

Christensen, Ken Ramshøj, Johannes Kizach, and Anne Mette Nyvad. 2013. Escape from the island: Grammaticality and (reduced) acceptability of wh-island violations in Danish. *Journal of Psycholinguistic Research* 42: 51–70. [CrossRef] [PubMed]

Cowart, Wayne, and Dana McDaniel. 2021. The That-trace Effect. In *The Cambridge Handbook of Experimental Syntax (Cambridge Handbooks in Language and Linguistics)*. Edited by Grant Goodall. Cambridge: The Cambridge University Press, pp. 258–77.

Crawford, Jean. 2012. Using syntactic satiation effects to investigate subject islands. Paper presented at the 29th West Coast Conference on Formal Linguistics, Tucson, AZ, USA, April 24; Edited by Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz and Alex Trueman. Somerville: Cascadilla Proceedings Project, pp. 38–45.

Do, Monica L., and Elsi Kaiser. 2017. The relationship between syntactic satiation and syntactic priming: A first look. *Frontiers in Psychology* 8: 1851. [CrossRef] [PubMed]

Featherston, Sam. 2021. Response methods in acceptability experiments. In *The Cambridge Handbook of Experimental Syntax (Cambridge Handbooks in Language and Linguistics)*. Edited by Grant Goodall. Cambridge: The Cambridge University Press, pp. 39–61.

Francom, Jerid Cole. 2009. Experimental Syntax: Exploring the Effect of Repeated Exposure to Anomalous Syntactic Structure–Evidence from Rating and Reading Tasks. Doctoral dissertation, University of Arizona, Tucson, AZ, USA.

Goodall, Grant. 2011. Syntactic satiation and the inversion effect in English and Spanish *wh*-questions. *Syntax* 14: 29–47. [CrossRef]

Hiramatsu, Kazuko. 2000. Accessing Linguistic Competence: Evidence from Children's and Adults' Acceptability Judgments. Doctoral dissertation, University of Connecticut, Storrs, CT, USA.

Hofmeister, Philip, T. Florian Jaegger, Inbal Arnon, Ivan A. Sag, and Neal Snider. 2013. The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes* 28: 48–87. [CrossRef] [PubMed]

Maia, Marcus. 2013. Linguística experimental: Aferindo o curso temporal e a profundidade do processamento. *Revista de Estudos da Linguagem* 21: 9–42. [CrossRef]

Perlmutter, David M. 1968. Deep and Surface Structure Constraints in Syntax. Doctoral dissertation, MIT, Cambridge, MA, USA.

Rosenbaum, Peter S. 1967. *The Grammar of English Predicate Complement Constructions*. Cambridge: The MIT Press.

Ross, John Robert. 1967. Constraints on Variables in Syntax. Doctoral dissertation, MIT, Cambridge, MA, USA.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Snyder, William. 1994. A psycholinguistic investigation of weak crossover, islands, and syntactic satiation effects: Implications for distinguishing competence from performance. Paper presented at the CUNY Human Sentence Processing Conference, CUNY Graduate Center, New York, NY, USA, March 17.

Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31: 575–82. [CrossRef]

Snyder, William. 2021. Satiation. In *The Cambridge Handbook of Experimental Syntax (Cambridge Handbooks in Language and Linguistics)*. Edited by Grant Goodall. Cambridge: The Cambridge University Press, pp. 154–80.

Sobin, Nicholas. 1987. The variable status of Comp-trace phenomena. *Natural Language & Linguistic Theory* 5: 33–60.

Sprouse, Jon. 2009. Revisiting satiation: Evidence for an equalization response strategy. *Linguistic Inquiry* 40: 329–41. [CrossRef]

Sprouse, Jon, and Sandra Villalta. 2021. Island effects. In *The Cambridge Handbook of Experimental Syntax (Cambridge Handbooks in Language and Linguistics)*. Edited by Grant Goodall. Cambridge: The Cambridge University Press, pp. 227–57.

Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax* 10: 80–126. [CrossRef]

languages

MDPI

Article

Sources of Discreteness and Gradience in Island Effects

Rui P. Chaves

Linguistics Department, University at Buffalo, Buffalo, NY 14260-1030, USA; rchaves@buffalo.edu

Abstract: This paper provides an overview of categorical and gradient effects in islands, with a focus on English, and argues that most islands are gradient. In some cases, the island is circumvented by the construction type in which the extraction takes place, and there is growing evidence that the critical factor is pragmatic in nature, contrary to classic and categorical accounts of island effects that are favored in generative circles to this day. In other cases, the island effect is malleable and can weaken with increased exposure to the extraction pattern, a phenomenon traditionally referred to as 'syntactic satiation'. However, it is not clear what satiation consists of. Some argue that it is nothing more than task adaptation (mere increased familiarity with the experimental paradigm, impacting difficult sentences more than easy ones), whereas others propose that it consists of a form of error-driven structure-dependent form of learning. The present paper discusses this controversy, and the broader adaptation debate, and argues that both task adaptation and grammatical adaptation are taking place during the processing of complex sentences, and that both frequency and attention are plausible factors to stimulate adaptation.

Keywords: Islands; satiation; frequency; adaptation

Citation: Chaves, Rui P. 2022. Sources of Discreteness and Gradience in Island Effects. *Languages* 7: 245. https://doi.org/10.3390/languages7040245

Academic Editors: Anne Mette Nyvad and Ken Ramshøj Christensen

Received: 4 March 2022
Accepted: 7 September 2022
Published: 21 September 2022

1. Introduction

There is growing evidence that repeated exposure to infrequent syntactic structures can lead to adaptation, as measured in faster reading times and/or increased acceptability. For example, certain illicit *wh*-movement structures known as 'islands' (Ross 1967) can become more acceptable, and are processed faster, with increased repeated exposure, a phenomenon often referred to as *syntactic satiation* (Snyder 1994, 2000; Stromswold 1986) The precise nature of syntactic satiation is not known. It could be an instance of *task adaptation* (Heathcote et al. 2000) (i.e., mere increased familiarity with the experimental paradigm, perhaps impacting difficult sentences more than easy ones as argued by Prasad and Linzen (2021)). Alternatively, it could be *syntactic adaptation* (Chang et al. 2006, 2012; Fine et al. 2010, 2013; Fine and Jaeger 2013; Sikos et al. 2016) (an error-driven structure-dependent form of statistical learning, whereby unexpected structures cause the processor to adapt to the contingencies of the input), or a combination of the two. Such changes in behavior are important because they can shed light on whether grammar is gradient, and fundamentally probabilistic, or categorical after all. This in turn is connected to broader questions about how language changes, and how it is learned by children as well as adults.

In Section 2 I provide an overview of the evidence suggesting that there are two major kinds of island phenomena. One the one hand we have categorial effects, which are due to some strict (syntactic or semantic) grammatical constraint, and in the other we have gradient effects, which are to a large extent caused by contextual or expectation-based factors. In some islands, there is a confluence of both types of phenomena, which are difficult to disentangle. In Section 3 I turn to amelioration effects caused by repeated exposure, which is a selective phenomenon, as certain island violations are more susceptible than others to ameliorate than others. Several kinds of account for this effect are surveyed, and it is argued that Brown et al. (2021) are incorrect in regarding all satiation as a form of task adaptation having nothing to do with grammar or island phenomena. To further disentangle task adaptation from syntactic adaptation, I describe a self-paced reading

experiment, using a classic garden-path effect, to show that increased reward leads to more adaptive behaviour in the critical region. The experiment suggests that fine-grained error-driven learning is taking place, and that frequency can compound with reward to and speed up sentence processing of complex sentences, over and above task adaptation.

2. Discreteness and Gradience

It has become increasingly clear that island effects are not created equal, and lie on a continuum constrained by multiple factors. At one end, we have islands that are categorical and exceptionless. These are construction-invariant (i.e., are active in any construction that involves unbounded extraction), immune to any form of principled circumvention (e.g., parasitism), insensitive to contextualization, and do not weaken with repeated exposure (i.e., satiation).[1]

A good example of how disparate island phenomena can be is the Coordinate Structure Constraint (Ross 1967), which is composed of two separate parts. One bans extraction *from* conjuncts, called the Element Constraint (Grosu 1973), and the other one bans extraction *of* conjuncts, named the Conjunct Constraint (Grosu 1973). There is good evidence that the two constraints are due to fundamentally different factors. Let us focus on the Conjunct Constraint first, illustrated in (1). This constraint is construction-invariant, since it arises in any kind of filler-gap dependency construction, be it interrogative, declarative or subordinate.

(1) a. *Who did you see Robin and _ yesterday?
b. *Who did you see _ and Robin yesterday?
c. *It was Alex who you saw Robin and _ yesterday.
d. *It was Alex who you saw _ and Robin yesterday.
e. *The person who you saw _ and Robin yesterday was Alex.
f. *The person who you saw Robin and _ yesterday was Alex.

All of the sentences in (1) become acceptable if the conjunction 'and' is replaced with a comitative like 'with', which serves to indicate that it is the coordination that hampers extraction. To my knowledge, nothing can improve the acceptability of Conjunct Constraint violations. This includes Across-the-Board (ATB) extraction, as in (2).[2]

(2) a. *Who did you see _ and _ yesterday?
b. *It was Alex who you saw _ and _ yesterday.

The insensitivity to ATB extraction is noteworthy because ATB extraction circumvents the part of the Coordinate Structure Constraint that bans extraction from conjuncts, the Element Constraint. This is illustrated in (3).

(3) a. *Who did you say Alex dislikes Robin and Mia absolutely loves _ ?
b. *Who did you say Alex dislikes _ and Mia absolutely loves Robin?
c. Who did you say Alex dislikes _ and Mia absolutely loves _ ?

What is more, filler-gap dependencies like (3a,b) can become more acceptable if the conjunction is interpreted asymmetrically (Kehler 2002; Lakoff 1986; Na and Huck 1992), as illustrated in (4). Here, the order of the conjuncts matters for the interpretation. For example, in (4a) the first conjunct is a preparatory action for the second conjunct, which expresses the main assertion. In (4b) the second conjunct is a consequence of the first, and in (4) we have a more complex case of the same kind of pattern. No such meaning-based amelioration can salvage Conjunct Constraint violations.

(4) a. Who did Sam pick up the phone and call _ ?

b. How much can you drink _ and still stay sober?

c. What did Harry buy _ , come home, and devour _ in thirty seconds?

Taken together, the foregoing data tell us that the Conjunct Constraint and the Element Constraint are due to very different factors. The former constraint is brought about by coordination itself (conjuncts cannot be extracted), which can be explained if conjunctions are markers that attach to heads rather than heads that select arguments (Abeillé and Chaves 2021; Chaves 2007). The Element Constraint, in contrast, seems to be caused by the symmetric interpretation of coordination, which can be predicted by independently motivated pragmatic factors; see Kehler (2002, chp. 5) and Kubota and Lee (2015) for a more detailed discussion.

Another island type that resists any form of amelioration is the Left Branch Condition (LBC). This prohibits the extraction of determiner expressions in languages like English, as seen in (5). Nothing can ameliorate the effect, including repeated exposure (Francom 2009; Goodall 2011; Hiramatsu 2000; Snyder 2000, 2017; Sprouse 2009, 2007).

(5) a. *Whose did you meet _ friend?
(cf. 'Whose friend did you meet _ ?')

b. *Which did you buy _ book?
(cf. 'Which book did you buy?')

c. *It was Robin's I liked _ painting the most.
(cf. 'It was Robin's painting I liked _ the most.)

Since English LBC effects appear in any construction (relative clauses, declaratives, and interrogatives), and are not subject to contextual amelioration of any kind, they are a good candidate for a syntactic constraint on extraction.

Languages that apparently allow LBC violations, like most Slavic languages, don't have determiners (Uriagereka 1988, p. 113), and therefore the extracted phrase is in apposition to the nominal head. No LBC violation occurs. This is best illustrated by languages, like French, that obey the LBC but have a special construction in which such extractions are apparently possible (Corver 2014). Consider the contrast illustrated by (6a,b).

(6) a. *Quels avez-vous acheté _ livres?
how-many have-you bought books

'How many books have you bought'

b. Combien a-t-il vendu _ de livres?
how-many has-he sold of books

'How many books did he sell?'

There are good empirical reasons to believe that there is no LBC violation in (6b). The phrase *de livres* is a post-verbal NP in French, and *combien* behaves more like a nominal than a canonical quantifier (Abeillé et al. 2004; Kayne 1981), since the former can appear without the latter in the presence of other licensors, including the preposition *sans* ('without') or negation, e.g., *Paul n'a pas lu* [*de livres*] ('Paul did not read any books'). If *combien* and the *de*-phrase are autonomous, then that means that no LBC violation occurs in (6b). I suspect something analogous occurs in Slavic languages.

There are other construction-invariant and categorical island effects, to be sure, such as the Preposition Stranding Ban in most languages that have prepositions, with the exception of some Germanic languages (including Danish, Dutch, English, Frisian, Norwegian and Swedish), as well as Berber, Hungarian, and Zoque (Emonds and Faarlund 2014, pp. 84–96).

At the other end of the spectrum we have island effects that are construction specific (i.e., are only active in certain types of unbounded dependency construction), permit systematic circumvention, exhibit varying degrees of acceptability depending on the exact wording (e.g., the plausibility of the content expressed, parsing difficulty caused by lexical ambiguity, garden paths, infrequent words, and/or stylistic issue), and can weaken with repeated exposure. According to the survey in Chaves and Putnam (2021), this is the true of the majority of known island effects; cf. with Szabolcsi and Lohndal (2017). In what follows I provide a brief overview of a number of island effects which are graded, malleable, and construction-specific.

2.1. Subject Islands

Subject Island violations, like the one in (7a), famously vanish with the presence of a second gap (Engdahl 1983) as illustrated by (7b), but see Chaves and Dery (2019) for concerns about such a paradigm.

(7) a. *Who did [the opponents of _] assassinate Castro?
b. Who did [the opponents of _] assassinate _ ?

The standard view that the second gap rescues the first by virtue of being outside the island is dubious, as Levine and Sag (2003), Levine and Hukari (2006, p. 256), and Culicover (2013, p. 161) note, because of examples like (8) in which both gaps are Subject Island violations. Such constructions should be completely ungrammatical.

(8) This is a man who [friends of _] think that [enemies of _] are everywhere.

More recently it has also been show that Subject Island effects can vanish if the extraction is from a relative clause, as in (9), which are attestations found by Culicover and Winkler (2022); see also Abeillé et al. (2020) for supporting experimental evidence.

(9) a. There are some things which$_i$ [fighting against $\__i$] is not worth the effort. Concentrating on things which can create significant positive change is much more fruitful.
[https://news.ycombinator.com/item?id=13946026, accessed on 7 January 2020]
b. I'm looking for someone who I click with. You know, the type of person who$_i$ [spending time with $\__i$] is effortless.
[https://3-instant.okcupid.com/profile/mpredds, accessed on 7 January 2020]
c. Survived by her children, Mae (Terry), Dale (Andelyn), Joanne (Gary), Cathy (Jordan), George, Betty (Tim), Danny (Angela); a proud grandmother of 14 grandchildren and 16 great-grandchildren, who$_i$ [spending time with $\__i$] was one of her finest joys;
[http://www.mahonefuneral.ca/obituaries/111846, accessed on 7 January 2020]

Attestations involving extraction from subject-embedded verbal structures are shown in (10). To my knowledge, there are no attested Subject Island violations that do not involve extraction from relative clause subjects.

(10) a. The eight dancers and their caller, Laurie Schmidt, make up the Farmall Promenade of nearby Nemaha, a town$_i$ that [to describe $\__i$ as tiny] would be to overstate its size.
(Huddleston et al. 2002, pp. 1094, ft.27)
b. In his bedroom, which$_i$ [to describe $\__i$ as small] would be a gross understatement, he has an audio studio setup.
(Chaves 2012, p. 471)

c. Leaving the room, she is quick to offer you some Arabic coffee and dates which$_i$ [to refuse _$_i$] would be insane because both are delicious, and an opportunity to relax and eat is welcome when working twelve hours. [www.thesandyshamrock.com/being-an-rt-in-saudi-arabia/, accessed on 7 January 2020]

Still, various authors such as Ross (1967, p. 242), Kluender (1998, p. 268), Hofmeister and Sag (2010, p. 370), Sauerland and Elbourne (2002, p. 304), Jiménez–Fernández (2009, p. 111), and Chomsky (2008, pp. 160, ft. 39), among others, have noted that slight rewording can attenuate Subject Island effects in interrogatives, as (11) illustrates.

(11) a. What did [the attempt to find _] end in failure?
(Hofmeister and Sag 2010, p. 370)

b. Which President would [the impeachment of _] not shock most people?
(Chaves and Putnam 2021, p. 80)

c. Which problem will [a solution to _] never be found?
(Chaves and Dery 2014)

Interrogative Subject Island violations like the above sometimes ameliorate with repetition (Chaves and Dery 2014; Clausen 2011; Francom 2009; Goodall 2011; Hiramatsu 2000; Lu et al. 2021). According to Chaves and Dery (2019), extractions from subjects like (12a) are initially less acceptable than their object counterparts in (12b), but as the experiment progressed the former became more acceptable, and by 12 exposures the two types of extraction were equally acceptable. This was replicated by Chaves and Putnam (2021, p. 213).

(12) a. Which stock does [the value of _] often parallels the price of the dollar?

b. Which stock does the value of the dollar often parallels [the price of _]?

The authors ensured that the acceptability differences in (12) were due to extraction (rather than to lexical biases, semantic plausibility, complexity, pragmatics, etc.), by making sure that their declarative counterparts shown in (13) were truth-conditionally near synonymous and expressed highly plausible propositions to begin with. This was done via a sentence acceptability norming experiment, with different participants.

(13) a. The value of this stock often parallels the price of the dollar

b. The value of the dollar often parallel the price of this stock.

Since the items expressed essentially the same proposition, this design avoided the concern raised by Kim (2021) about the factorial design adopted by Sprouse (2007), which does not control for important non-syntactic factors and therefore has limited ability to identify the exact nature of island effects. Chaves and Dery (2019) also compared acceptability and the online processing of near-synonymous sentence pairs like (12), which express essentially the same proposition. Any acceptability differences must come from the extraction itself.

The fact that no such dramatic acceptability increase was observed in the ungrammatical controls (including in a later replication by the same authors) suggests that Subject Island effects can vanish, in ideal conditions. That is, if the items are not too complex, express highly plausible propositions, and participants are sufficiently exposed to such structures. A similar effect was also observed in terms of reduced reading times around the gap site on a subsequent experiment in Chaves and Dery (2014). In other words, speakers can adjust to unusual gaps and the associated semantic-pragmatic consequences. The asymmetry between subject and object subextraction is not categorical, and can be countered in ideal conditions.

The conclusion is that English Subject Islands are most likely not purely syntactic phenomena. The effect is not present in relative clauses, and is sometimes graded elsewhere. But what, then, is behind such otherwise strong islands? One possibility is that extraction from subject phrases is dispreferred when the subject is expected to be discourse-old. Subject phrases are used typically used for topic continuity (Chafe 1994; Kuno 1972; Lambrecht 1994). For example, subject phrases are more likely to be pronominal or elliptical than objects (Michaelis and Francis 2007). Consequently, there is a conflict between the discourse function of the extracted element (focus) and the discourse function of the subject phrase itself (Abeillé et al. 2020; Erteschik-Shir 2006b; Goldberg 2006; Takami 1988; Van Valin 1986). Extracting from a discourse-old subject not only involves a structurally unexpected move, so to speak, but also contextually unusual state-of-affairs, one in which a discourse-old referent is linked to a subordinate referent that can be the focus. No such contradiction arises in relative clauses, because their subjects are under no obligation to be a main topic or focus.

According to Kluender (2004, p. 495), 'Subject Island effects seem to be weaker when the *wh*-phrase maintains a pragmatic association not only with the gap, but also with the main clause predicate, such that the filler-gap dependency into the subject position is construed as of some relevance to the main assertion of the sentence'. In other words, the subject-embedded referent must contribute to the interpretation of the main predication. For example, in (11a) the extraction is licit because whether or not the attempt to find x ends in failure crucially depends on the identity of x; the search failed precisely because of the nature of what was sought. Similarly, whether or not an impeachment shocks most people crucially depends on the one that is impeached, and whether or not a solution is found crucially depends on the identity of the problem.

Chaves and Putnam (2021, p. 228) found supporting experimental evidence for such a relevance constraint. A total of 20 experimental items were constructed, each of which had two versions, as seen in (14). The extracted referents in the –Relevant condition are less important for the situation described by the sentence as compared to the items in the +Relevant condition.

(14) a. Which joke was the punchline of extremely offensive?
(+Relevant)

b. Which joke was the punchline of overheard by the teacher?
(–Relevant)

To ensure that the +Relevant condition items were indeed more relevant than the –Relevant condition items, a norming experiment in which a different group of participants were asked to use a 5-point Likert scale to state to what extent they agreed with statements like (15), created from the 20 original experimental items.

(15) Whether the punchline of a joke is [offensive / overheard by the teacher] depends on what the joke is.

To ensure that any difference in acceptability between the item pairs was due to extraction and not to semantic or pragmatic differences between the item pairs, a norming experiment was conducted to measure the acceptability of the declarative counterparts of the 20 items, illustrated in (16). The goal of this task is to ensure that the non-extracted counterparts of the items were equally acceptable to begin with.

(16) The punchline of this joke was extremely offensive/overheard by the teacher.

After these norming experiments, acceptability ratings were collected for the 20 Subject Island items like (14). A Cumulative Bayesian Linear Regression model with sentence acceptability ratings as a dependent variable and the mean relevance ratings per item from the questionnaire experiment as the independent variable (allowing for the intercept to

vary with items and declarative acceptability ratings as random effects) found a significant effect ($\beta = 0.08$, $SD = 0$, CI = [0.07,0.08], P($\beta > 0$) = 1). These results suggest that the more important the extracted referent is for the proposition described by the utterance, the more acceptable the subject subextraction. This is consistent with the view in which not all subject embedded referents are equally biased to be assigned the same pragmatic function as the subject referent. This depends on the predication, the proposition, and the context. Moreover, whether or not a referent embedded in a discourse-old subject is interpreted as new and has an impact on the main predication is a matter of degree, and therefore it is not surprising that with repeated exposure such constructions sometimes become more and more acceptable. To conclude, Subject Islands are not construction-invariant, and even when they are active, their effect is gradient. Although a syntactic account may be possible, stipulating that in certain constructions extraction is allowed, it is unclear how such an account can *explain* why things are the way they are on independently motivated grounds.

2.2. Adjunct Islands

A similar situation arises in connection with Adjunct Islands. First, they can be circumvented by the presence of a secondary gap (Engdahl 1983), as illustrated by (17) and (18). But these sentence pairs have radically different meanings, and therefore it is not clear in which sense the main gap can be said to rescue the secondary gap. Indeed, it is well-known that such environments are not categorical boundaries to extraction, given examples like (19).

(17) a. *Which printouts$_i$ did Kim discard thumbtacks [without reading $\__i$]?
b. Which printouts$_i$ did Kim discard $\__i$ [without reading $\__i$]?

(18) a. *Which colleague$_i$ did John slander Robin [because he despises $\__i$]?
b. Which colleague$_i$ did John slander $\__i$ [because he despises $\__i$]?

(19) a. Who$_i$ did you go to Girona [in order to meet $\__i$]? (Hegarty 1990)
b. What$_i$ do you think Robin computed the answer [with $\__i$]? (Bouma et al. 2001, p. 45)
c. Which movies$_i$ does Sean Bean die [in $\__i$]? (Chaves and Putnam 2021, p. 87)
d. Which temperature$_i$ should I wash my jeans [at $\__i$]? (Chaves 2013)

There is no independently motivated empirical reason to assume that these adjuncts combine with their VP heads in different ways (Truswell 2011), which suggests that syntax is not the source of the island effect. Müller (2017) provides sentence acceptability evidence from Swedish suggesting that extraction from tensed adjuncts is contingent on the degree of semantic-pragmatic cohesion between the matrix, and similar results are reported for Norwegian by Bondevik (2018). More recently, Kohrt et al. (2020) and various others show that semantic factors play a critical role in English Adjunct Islands.

As in the case of Subject Islands, clausal Adjunct Island violations are usually stronger than phrasal violations. Compare (17a) and (18a) with (20).

(20) a. *Who$_i$ did John come back before I had a chance to talk to $\__i$? (Huang 1982, p. 491)
b. *Who$_i$ did Mary cry after John hit $\__i$? (Huang 1982, p. 503)

But Gibson et al. (2021) recently show that if a supporting context is provided, then island effects in tensed adjuncts is significantly ameliorated, suggesting that pragmatics plays a

role as well. Further evidence for the presence of semantic-pragmatic factors comes from the fact that the most acceptable Tensed Adjunct Island violations involve relative clauses which express assertions rather than backgrounded information. This is illustrated in (21).

(21) a. I got to do things$_i$ in the film that, if you did $\_\,_i$ on the street they'd send you away Epstein (2013).
(Chaves and Putnam 2021, p. 91)

b. I called the client who$_i$ the secretary worries if the lawyer insults $\_\,_i$.
(Sprouse et al. 2016)

c. This is the watch$_i$ that I got upset [when I lost $\_\,_i$].
(attributed to Ivan A. Sag (p.c.) by Truswell (2011, pp. 175, ft.1))

Indeed, Sprouse et al. (2016) found evidence of an Adjunct Island effect in interrogatives but no such effect in relative clauses like (21b). See also (Kush et al. 2018, 2019), and Müller and Eggers (2022) for similar findings about such relatives in English and other languages. In sum, Adjunct Islands are not construction-invariant, and seem to be sensitive to semantic and pragmatic factors (Kohrt et al. 2018a; Kohrt et al. 2018b; Müller and Eggers 2022). The parallel with Subject Islands does not stop here. Repeated exposure to interrogative Adjunct Islands can lead to amelioration effects (Chaves and Putnam 2021, pp. 232, 238). This includes clausal islands like (22), which by the end of the experiment were as acceptable as grammatical controls.

(22) Who would Amy be really happy [if she could speak to _]?

2.3. Complex NP Constraint

There are various other island effects that are similarly not construction-invariant, and which are attenuated when extraction occurs from structures that do not express backgrounded content. Complex NP Constraint (CNPC) phenomena, illustrated in (23) in one such case.

(23) a. *Who$_i$ did Kim believe [the claim that Robin saw $\_\,_i$]?

b. *Which student$_i$ should we report [the teacher who punished $\_\,_i$]?

These island are graded, as has long been noted (Culicover 1999; Deane 1991; Erteschik-Shir and Lappin 1979; Kluender 1998; Kuno 1987). Compare (23b) with the isomorphic example in (24). Furthermore, CNPC violations ameliorate with repeated exposure, as shown by Snyder (2000), and Goodall (2011).

Erteschik-Shir (1977, chp. 2) first noted that in CNPC exceptions the matrix predicate is in general less informative than the embedded, and main verbs like *hear* and *know* are almost devoid of semantics, which makes it more likely for the main action to be conveyed by the subordinate clause. See Vincent (2021) and Vincent et al. (2022) for experimental evidence confirming that English should be counted among the languages that allow extraction from relative clauses in environments such as the one in (24).

(24) Which kid$_i$ did you hear [a rumor that my dog bit $\_\,_i$]?
(Chaves and Putnam 2021, p. 67)

This also explains why CNPC effects tend to vanish in relative clauses that express the assertion of the utterance, as in (25). See Erteschik-Shir and Lappin (1979), McCawley (1981, p. 108), Chung and McCloskey (1983) for more examples, and Kush et al. (2013) and Sprouse et al. (2016) for supporting experimental evidence. The situation is not unlike that of Subject and Adjunct Islands.

(25) a. Which diamond ring$_i$ did you say there was [nobody in the world who could buy _$_i$]?
(Pollard and Sag 1994, p. 206)

b. John is the sort of guy$_i$ that I don't know [a lot of people who think well of _$_i$].
(Culicover 1999, p. 230)

2.4. Factive Islands

Factive Island phenomena exhibit various kinds of circumvention phenomena. As Szabolcsi and Zwarts (1993) originally noted, when a question necessarily has a unique true (and non-negative) answer then the presence of a factive verb hampers extraction, as illustrated in (26). See Oshima (2007), Schwarz and Simonenko (2018), and Abrusán (2014) for elaborations of this conclusion.

(26) a. #Who did Robin know that [Alex helped _ first]?

b. Who did Robin say that [Alex helped _ first]?

As a consequence, there are two ways to circumvent the effect in (26a). One way is to make the question not have a necessarily unique true answer, which can be achieved by replacing the one-time adverb *first* with any other kind of adverb:

(27) Who did Robin know that [Alex helped _ yesterday]?

The other way to circumvent the effect is to convert the unbounded dependency to a declarative, as in (28). This means that Factive islands are not construction-invariant, since they disappear in non-interrogative extractions.

(28) a. It was Kim who Robin knew that [Alex helped _].

b. I met the person who Robin knew that [Alex helped _].

c. KIM, Robin knew that [Alex helped _]. MIA he didn't.

But there are other, more subtle, island effects in clausal complements, illustrated in (29), where the interrogatives are not required to have a unique true answer. Here, it is the mere presence of a factive or manner-of-speaking verb that hampers extraction.

(29) a. What did John say that Mary bought _ ? (Bridge verb)

b. ??What did John know that Mary bought _ ? (Factive verb)

c. ??What did John whisper that Mary bought _ ? (Manner-of-speaking verb)

Most researchers seem to agree that the explanation for these for Bridge verb effects is at least in part pragmatic, although they disagree in the details (Ambridge and Goldberg 2008; Erteschik-Shir 2006a; Kothari 2008; Liu et al. 2022), and if Tonhauser et al. (2018) and Degen and Tonhauser (2022) are correct about factivity being a matter of degree, this would explain why such island effects are fuzzy.

For example, Ambridge and Goldberg (2008) provide evidence suggesting a pragmatic explanation: the more backgrounded the proposition, the stronger the island effect. Liu et al. (2019) challenge these findings and instead provide evidence suggesting that the frequency with which verbs are used in the clausal complement frame is responsible for acceptability contrasts observed by extracting from factive and manner clausal complements. Liu et al. (2022) conjectures that discourse, semantic, and structural factors might conspire to give rise to the observed frequency distributions, which in turn give rise to acceptability ratings.[3]

2.5. Interim Summary

Most of the islands discussed above are not construction-invariant. They are stronger in interrogatives than in relative clauses that express assertions, for example. This suggest a common thread between the Element Constraint, Subject Islands, Adjunct Islands, Factive Islands, and the Complex NP Constraint: asserted content more readily allows extraction than backgrounded (non-at-issue content); cf. with Erteschik-Shir and Lappin (1979), Kuno (1987), Goldberg (2013), Chaves and Dery (2019), and Abeillé et al. (2020).

This observation allows us to make further predictions. For example, it means that extraction from parentheticals should be impossible, regardless of the construction, because parentheticals by definition express suppletive information, orthogonal to the main assertion. This prediction is borne out in the contrasts in (30) and (31).

(30) a. The union leaders–in case you missed that article–refused to sign the contract.

b. *It was that article that the union leaders–in case you missed _ –refused to sign the contract.

c. *What the union leaders –in case you missed _ –refused to sign the contract was that article.

(31) a. David Johnson – I am not sure Robin told you this–refused to sign the contract.

b. *It was Robin who David Johnson–I am not sure _ told you this –refused to sign the contract.

c. *Who David Johnson–I am not sure _ told you this–refused to sign the contract was Robin.

Why are island effects gradient, even in interrogative environments? Tonhauser et al. (2018) provides evidence that whether or not speakers commit to the content expressed by subordinate clauses is a matter of degree, as it depends on a number of factors, including the prior probability of the event that is described. If this is correct, then it would provide an explanation for why *wh*-phrases embedded in the subjects of certain interrogatives are more readily interpreted as Foci than others, i.e., more readily extracted, and so on. Another possibility is that the increase in acceptability is due to more general factors, independent of islands, which have more to do with how informants adapt to psycholinguistic tasks. I turn to this matter in the following section.

3. Satiation

As discussed above, even when the filler-gap construction type is island-inducing, it is often the case that the island effect can be attenuated with repeated exposure (Chaves and Dery 2014, 2019; Clausen 2011; Do and Kaiser 2017; Francom 2009; Goodall 2011; Hiramatsu 2000; Hofmeister 2015; Lu et al. 2021; Snyder 2000, 2017) as discussed above. To be sure, such amelioration is not consistently observed, suggesting that different results arise because different researchers have used different stimuli and different exposure rates (Chaves and Dery 2019; Hiramatsu 2000; Hofmeister 2011; Snyder 2017). In particular, the role of stimuli design cannot be overstated. If sentences that are too complex or awkward are used, satiation is less likely to be observed (Hofmeister 2011; Hofmeister and Sag 2010). Consider for example the sample of items in (32), from Sprouse et al. (2012).

(32) a. *What$_i$ do you faint if the actors forget $\__i$ on stage?

b. *What$_i$ do you sneeze if the dog owner leaves open $\__i$ at night?

c. *What$_i$ do you cough if the tourists photograph $\__i$ in the exhibit?

d. *What$_i$ do you laugh if the heiress buys $\__i$ at the auction?
(Sprouse et al. 2012)

Now contrast these with (33), which Chaves and Putnam (2021, p. 238) found induce satiation. Crucially, the adjunct clause coheres much better with the main predication because it expresses a cause that triggers the state described by the psychological predicate. In contrast, there is no obvious relation between the main predication and the conditional clause in (32).

(33) a. Who$_i$ would Amy be really happy if she could speak to $\__i$?
b. What$_i$ would Jill get really angry if she missed $\__i$?
c. What$_i$ would Allison be really upset if she forgot $\__i$?

The low acceptability of such tensed Adjunct Island violations and their lack of satiation is likely to be due, at least to some extent, to the described propositions. For instance, people don't routinely faint when something is forgotten on stage as in (32a), or typically sneeze if dog owners leave something open at night, as in (32b). These are perfectly possible propositions, but they describe rather unusual situations. The event described by the matrix predication does not cohere particularly well with that of the adjunct's predication. In order to avoid this kind of problem, one would have to norm the declarative counterpart of these items, to ensure that all are equally felicitous and plausible.

The amelioration effect caused by repeated exposure is referred to as *syntactic satiation*, in analogy to the phenomenon of semantic satiation, whereby repetition causes a word or phrase to temporarily lose meaning for the listener. There are two problems with this terminology. First, it is perfectly reasonable that the increase of acceptability is caused by semantic and pragmatic factors, over and above syntactic factors. Second, whereas semantic satiation is a fairly well-understood general reactive inhibition phenomenon (a bottom-up processing process associated with lower level neural mechanisms of inhibition), the increase of acceptability during sentence processing is selective: certain island violations robustly ameliorate with repetition, whereas others simply do not, as discussed above. In contrast, repetition of any lexical item can induce the semantic satiation effect. Syntactic satiation seems to be faciliatory in nature, rather than inhibitory, because repeated exposure to island violations does not lead to loss of sentence meaning. Thus, comprehension question accuracy does not decline as island effects ameliorate.

A more concerning problem is that it remains unclear what syntactic satiation actually amounts to. It could be a form of adaptation, caused by changes in the activation of the representations in declarative memory (i.e., a form of priming), residual activation (the mechanism that accounts for priming), a change in the procedural knowledge required to construct the relevant structures (adaptations to the parsing strategy), or belief-updating (violated expectations lead to probabilistic updates, under a Bayesian interpretation).

3.1. Adaptation

Sensory input is typically noisy and ambiguous, and individuals respond to the challenges created by such variation by using probabilistic expectations (Anderson 1990; Gigerenzer et al. 1999; Newell and Simon 1972). For example, infants already exhibit the ability to integrate prior beliefs, knowledge, and expectations about human actions with new evidence provided by the environment (Xu and Kushnir 2013), and use new evidence to modify their prior expectations (Brandone et al. 2014). Linguistic input is particularly noisy, ambiguous, and variable across individuals and contexts, and therefore it is expected that speakers can adapt to the contingencies of the input. This would enable individuals to make heuristic predictions and robustly cope with such a dynamic linguistic input. For example, it is known that comprehenders create expectations about upcoming words (Altmann and Kamide 1999; Arai and Keller 2013; Creel et al. 2008; DeLong et al. 2005; Kutas and Hillyard 1984; Metzing and Brennan 2003), about upcoming lexical categories (Gibson 2006; Levy and Keller 2013; Tabor et al. 1997), and about syntactic structures (Farmer et al. 2014; Fine et al. 2010, 2013; Fine and Jaeger 2013; Kamide and Mitchell 1997; Lau et al. 2006; Levy 2008; Levy et al. 2012; MacDonald et al. 1994; Malone and Mauner 2018; Stack et al.

2018; Staub and Clifton 2006; Wells et al. 2009), among other modalities of linguistic input. In what follows I will provide a brief survey of this literature and the controversy therein about the nature of adaptation. See Kaan and Chun (2018) for a detailed overview.

3.2. Adaptation in Garden-Path Sentences

Fine et al. (2010), Kamide and Mitchell (1997), Farmer et al. (2014) and others provide evidence suggesting that syntactic expectations are malleable and quickly adapt to changes in the input. Fine and Jaeger (2013) argue that repeated exposure to *a priori* unexpected structures can reduce, and even completely invert, their processing disadvantage, and *a priori* expected structures can become less expected (even eliciting garden paths) in environments where they are hardly ever observed. As illustrated in (34), past participle verb forms often give rise to a temporary ambiguity between a main verb parse like (34a), and a relative verb parse, seen in (34b). However, (34a) and (34b) differ in that the latter consistently elicits a garden-path effect, because the main verb use of *warned* is much more likely than the relative verb use according to corpora evidence (Roland et al. 2007). This effect has been detected by various researchers, including Stack et al. (2018), Malone and Mauner (2018), Prasad and Linzen (2019 2021), Dempsey et al. (2020).

(34) The experienced soldiers . . .
a. . . . warned about the dangers before the midnight raid.
b. . . . warned about the dangers conducted the midnight raid.

But by making the relative verb use more frequent than the main verb use in a controlled experiment, Fine and Jaeger (2013) found that the garden-path effect can flip: the relative verb parse becomes the default preferential parse, and the main verb parse becomes dispreferred. By the end of the experiment, sentences like (34b) no longer exhibit a garden-path effect because the relative verb parse is now the most frequent and preferential parse, whereas sentences like (34a) now yield a garden-path effect. The latter is called a *reverse ambiguity effect*. Fine et al. (2013) argue that comprehenders adapt to the statistics of the current linguistic environment by generating expectations that reflect the distribution of actual events in the environment. This rational strategy allows comprehenders to reduce the average prediction error experienced during processing.

More recently, Lu et al. (2021) provide evidence suggesting that comprehenders can exhibit speaker-specific satiation to Subject Islands, and argue that syntactic satiation in island phenomena is a form of Bayesian learning *a la* Fine et al. (2010).

The reverse ambiguity effect that penalizes a priori preferred structures found by Fine and Jaeger (2013) seems to be elusive, however. It was replicated by Sikos et al. (2016), but not by Stack et al. (2018). See also Jaeger et al. (2018) for a response. Now, it is worth pointing out that these studies used different experimental items, different numbers of participants, different amounts of exposure, different compensation levels for participants, and different statistical methods. As I will discuss below, at least some of these may play a crucial role in promoting adaptation.

Second, although the reading times of garden-path sentences decreased in all of the aforementioned studies, this also happened for all other sentences, including controls that were not temporarily ambiguous. In fact, there is independent evidence that reading times generally decrease exponentially as a function of practice (Heathcote et al. 2000). Given this evidence, Stack et al. (2018), Prasad and Linzen (2019, 2021), and Dempsey et al. (2020) argue that the reduction in reading time due to syntactic adaptation is confounded with a more general adaptive phenomenon, called *task adaptation*: adaptation driven instead by increased familiarity with the experimental procedure, rather than by syntactic structure. For Dempsey et al. (2020), task adaptation is what is commonly referred to as syntactic satiation. The latter does not directly depend on the syntactic structure of the sentence, and could be due to a number of factors, such as word frequency, plausibility, predictability, and syntactic disambiguation difficulty.

3.3. Syntactic Satiation as Adaptation in Islands

For Brown et al. (2021) syntactic satiation in islands is a form of task adaptation, and has nothing to do with grammar or island phenomena. In their experiments, only the items with intermediate acceptability became more acceptable, and they did so only at the beginning of the experimental session, regardless of syntactic construction. However, other island satiation studies find different patterns. For example, Hiramatsu (2000) found Subject Island satiation with 7 exposures but not with 5. This should be impossible if satiation mainly occurred at the beginning of the experiment. Similarly, Hofmeister (2015) found that Adjunct Islands satiate after 8 exposures but not before (this experiment replicated); see Chaves and Putnam (2021, p. 232) for details. None of these results are expected if satiation is mainly located at the beginning of the experiment.

What is more, different conditions usually satiate at different rates, contrary to the generalization put forth by Brown et al. (2021). In (35) are examples of item types used by Hofmeister (2015).

(35) a. Just a few years ago, Mosul was a city which terrorists would have thought twice before attacking.
[Adjunct Island condition]

b. Just a few years ago, terrorists would have thought twice before attacking the city of Mosul.
[Non-island condition]

c. The rebels in the jungle captured the diplomat who pleaded with the villagers after they threatened to kill his family for not complying with their demands.
[Right-branching]

d. The diplomat who the rebels who were in the jungle captured pleaded with the villagers after they threatened to kill his family for not complying with their demands.
[Center-embedded]

Linear Mixed-Effect models with acceptability as the dependent variable and the presentation order as the predictor (allowing the intercept to be adjusts by list and item, as random effects) reveal that the acceptability center-embedding condition increased significantly as the experiment progressed ($\beta = 0.02, SD = 0.005, t = 4.042, p < 0.0001$), as did the adjunct island condition ($\beta = 0.01, SD = 0.004, t = 3.89, p = 0.0001$), whereas the right-branching condition did not ($\beta = -0.01, SD = 0.005, t = -0.19, p = 0.84$). The non-island condition improved as well, but the effect size was much the smallest ($\beta = 0.007, SD = 0.002, t = 2.66, p = 0.007$). This is seen in Figure 1.

Crucially, the right-branching condition received middle ratings, and yet did not experience any acceptability changes. Moreover, adjunct island items (at the very bottom of the acceptability range) only rise sharply and consistently in the last two thirds of the experiment. These results are unexpected for Brown et al. (2021).

As a final example, consider the satiation patterns of three different types of clausal adjuncts from the data in Chaves and Putnam (2021, p. 238), shown in Figure 2. Again, extractions from one item type (in this case, conditional clauses like (22) above) exhibit a more consistent trajectory than the others. Again, these results challenge the generality of the conclusions of Brown et al. (2021).

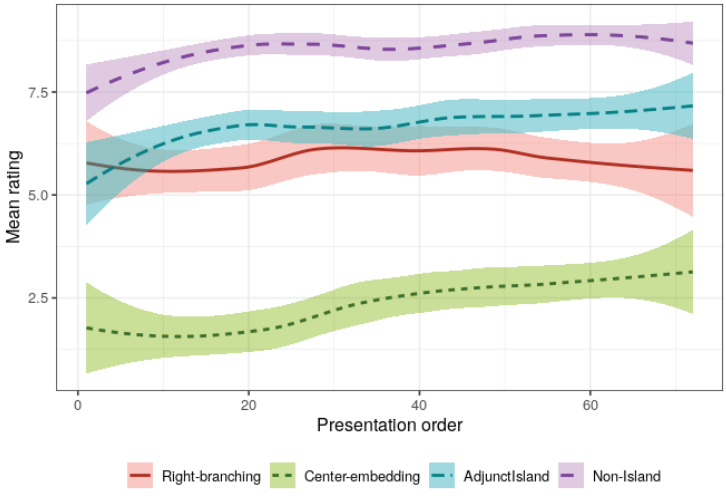


Figure 1. Differential effect of repeated exposure in Hofmeister (2015).

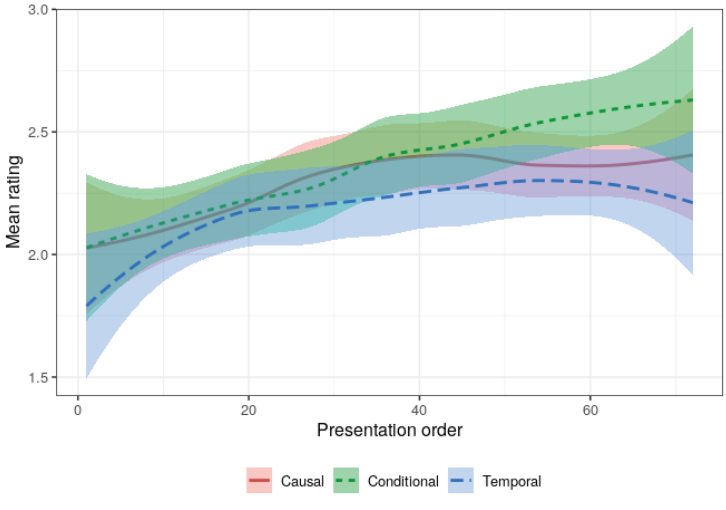


Figure 2. Effect of repeated exposure in Clausal Adjuncts in Chaves and Putnam (2021, p. 238).

Prasad and Linzen (2021) suggest that sentences that are difficult to process undergo a sharper rate of task adaptation than easier sentences, which in turn overwhelms the effect of syntactic adaptation, if any exists. They argue that the effect of syntactic adaptation is very small, and required very large numbers of participants (around 1000). If this is the case, then there should be a correlation between acceptability and the satiation coefficient. To test this hypothesis, data from three separate experiments was used. First, the clausal adjunct island satiation data mentioned above (Chaves and Putnam 2021, p. 238) were obtained and the (significant) satiation coefficients, per item, were compared with the respective mean acceptability ratings. The correlation was not significant ($t = 2.06$, $p = 0.13$), and had it been significant, it would have been positive, not negative. Next, the significant satiation coefficients from the Adjunct Island violations in Hofmeister (2011) were also computed, by item, as above, and correlated with the respective mean acceptability ratings. Again, no

significant correlation was found ($t = -0.28$, $p = 0.79$). The same was done for the Subject Island satiation data from Chaves and Dery (2019), Chaves and Putnam (2021, p. 212), and again the correlation was not significant ($t = 1.12$, $p = 0.34$). These results are the opposite of what Prasad and Linzen (2021) would predict if these island satiation effects affected low acceptability sentences more than high acceptability sentences. And if more extreme island violations do not satiate more, then island satiation cannot amount to just task adaptation, according to the logic of Prasad and Linzen (2021).

3.4. Disentangling Task Adaptation from Syntactic Adaptation

Malone and Mauner (2020) develop a new approach to decoupling syntactic adaptation from task adaptation, and show that the former is detectable without large numbers of participants, and robust. In a nutshell, they propose that the effect of task adaptation is dealt with by using *Task-Adapted Reading Times* (TART). These are conceptually similar to residual reading times that correct for the effect of word length on reading speed within individual participants. The TART procedure uses the speed-up in reading times in the distractor items as a proxy for task adaptation. The assumption is that as distractors are structurally unambiguous, uncomplicated sentences, any reduction in reading times over the course of the experiment should (i) not be due to syntactic adaptation, and (ii) be due to task adaptation, as participants mechanically or cognitively adjust to the task. Distractor regions 4 through 11 were selected, and regressed onto stimulus order (not critical item order) for each participant. Because these regressions do not include critical items, and no learning should occur in distractors, this method can measure task adaptation, unconfounded with syntactic adaptation.

As TART involves regressing reading times over distractor item order, the first step is to correct reading times in the selected region of analysis by residualizing reading times to correct for word length. The second step is to then regress item order over the length-corrected distractor reading times discussed above, with participant as a random factor. The result should be a model that captures the unique rate of increase in reading time over the course of the experiment for each participant. These TART values are then subtracted from the reading times of each of the critical item regions, and the resulting reading times are residualized based on word length, per region and participant, as is standard (Trueswell and Tanenhaus 1994). The new reading times, now adjusted for both character length and participants' unique increase over time due to irrelevant task adaptation, are now ready to be fit in the primary analytic regression model. If distractors are structurally diverse, unambiguous, and uncomplicated sentences, then all syntactic adaptation must come from the regularities in the critical items.

3.5. The Role of Reward

It is now well-known that learning requires attention, alertness, and focus, and that predicted reward (dopamine) can not only help engage these systems but also promote synaptic plasticity by enhancing long-term potentiation and depression (Legenstein et al. 2008; Otmakhova and Lisman 1996; Reynolds et al. 2001; Schultz 1998). It follows that adaptation in language processing should be sensitive to the predicted reward, not just to structural frequency and task adaptation. There is currently no standard for the compensation of participants in psycholinguisics experiments, and perhaps this is a problem. For example, Fine et al. (2010) compensated participants with course credit, Fine and Jaeger (2013) paid participants $10, Stack et al. (2018) paid $4, Dempsey et al. (2020) paid $3, and Prasad and Linzen (2021) paid $6.51 per hour. It is therefore possible that these participants experienced different levels of motivation and focus while performing this task, which had an effect on the probability of learning regularities in the items. As Christianson et al. (2022) show, both online and offline measures of processing and comprehension are susceptible to focus and motivation levels, leading to results that are not reflective of normal human language processing.

To probe for the effect of the predicted reward and provide independent support for the TART methodology, an experiment was designed and conducted to determine whether syntactic adaptation is sensitive to the predicted reward, over and above structural frequency. I focused on a garden-path effect rather than on an island because there is no question that such constructions are grammatical, and all of the literature on task adaptation has focused on garden-paths. Future work should probe island constructions.

4. TART Reward Experiment: Adaptation to Complex Sentences

4.1. Methods

Subjects

In this between-subjects experiment, 100 participants with US-based IP addresses were recruited via the Amazon Mechanical Turk marketplace, all of which self-reported as having grown up speaking English as a first language via a language questionnaire conducted after the experiment concluded. Participants were informed that their responses to the language questionnaire had no bearing on their compensation.

Only subjects with at least 98% approval rating from previous jobs and with over 10k previous tasks approved were allowed to participate. Participants were told the experiment consisted of reading 32 sentences and answering yes/no comprehension questions correctly. Participants were compensated with $2.4 for their participation.

Participants were randomly assigned to either the Control group or the Bonus group. All participants were informed that the experimenters might not be able to compensate them if their comprehension accuracy dipped significantly below 70%, although in practice no participants were excluded from compensation. The individuals from the Bonus group saw additional text and instructions informing them that if their comprehension question accuracy was above 75%, they would receive a bonus of $4.80, for a total of $7.2. The participants from each group saw the same stimuli.

Ethics statement

This study was conducted with the approval of the Institutional Review Board of the University of the State of New York at Buffalo. All participants gave their informed written consent.

Materials

A total of 16 items were constructed, all of which exhibited the classic subject/object ambiguity in (36), whereby a noun phrase (underlined) is a temporally plausible object of the preceding verb, but is in fact the subject of the following main verb (bold font) (Christianson et al. 2001; Ferreira and Henderson 1990; Frazier and Rayner 1982; Jacob and Felser 2016).[4] This late closure parse is well-known to be susceptible to priming, as reflected behaviorally by decreased reading times (Noppeney and Price 2004; Traxler 2015), and physiologically by attenuated responses in the left temporal pole (Noppeney and Price 2004).

(36) a. After $_1$| the $_2$| Mayor $_3$| visited $_4$| <u>the</u> $_5$| <u>patients</u> $_6$| **were** $_7$| moved $_8$| to $_9$| different $_{10}$| rooms. $_{11}$|
[The Mayor paid a visit after the patients were moved. True or False?]

b. While $_1$| the $_2$| customers $_3$| ate $_4$| <u>some</u> $_5$| <u>food</u> $_6$| **was** $_7$| cooking $_8$| on $_9$| the $_{10}$| grill. $_{11}$|
[The customers ate only after all the cooking was done. True or False?]

Half the items were disambiguated by 'was'; the other half by 'were'. The prepended adverbs were 'after', 'although', 'as', 'though', 'when' and 'while', evenly distributed across items. To maximize the garden path effect, the subordinate verbs came from a subset of verbs from Ferreira and Henderson (1991) and Staub (2007) that had the highest proportion of transitive uses relative to intransitive uses, according to both (Gahl et al. 2004) and to a corpus study using the Corpus of Contemporary American English (Davies 2008). The 16 items were pseudo-randomized and interspersed with 16 distractors, illustrated in (37). The

latter used the same prepended adverbs (plus the adverbs 'because', 'if' and 'whenever'), evenly distributed across distractors, and a variety of verbal structures different from the items. Across items and distractors, no two stimuli contained the same verb, as to avoid priming effects caused by verb repetition (Fine and Jaeger 2016; Traxler and Pickering 2005). Although all participants in the experiment saw the same stimuli, no two participants saw the same order of stimuli.

(37) a. Though $_1$| the $_2$| bus $_3$| driver $_4$| missed $_5$| a $_6$| street $_7$| Sue $_8$| was $_9$| at school $_{10}$| on time. $_{11}$|
[Sue brought a child home after the bus missed its stop. True or False?]

b. If $_1$| the $_2$| radar $_3$| is $_4$| correct $_5$| the $_6$| storm $_7$| will $_8$| be $_9$| here $_{10}$| tomorrow. $_{11}$|
[The radar data can be used to make predictions about the weather. True or False?]

Procedure

Subjects read sentences in a self-paced moving window display (Just et al. 1982), using the self-paced reading mode of the PCIbex platform (Zehr and Schwarz 2018). Three practice trials were conducted before the experiment proper started. All experimental items were followed by a Yes/No comprehension question probing the lingering of the initial interpretation. The form of the comprehension questions varied from item to item, to prevent participants from strategizing how to answer the comprehension questions. The correct answer was "yes" half of the time, and after submitting each answer participants were informed about whether their selection was correct or not. The stimuli were pseudo-randomized so that no two participants saw the items in the same order and no more than two critical items were allowed to immediately follow each other. Participants took an average 15 min to complete the experiment, meaning that Control group participants were paid at an hourly rate of about $9.4 while the Bonus group participants were paid at a $28.8 hourly rate.

Filtering

Participants with comprehension question accuracy below 75% were excluded, resulting in 12% of data loss (11.3% data loss for the Control group, and 11.2% for the Bonus group). Only distractors were used for this participant exclusion criterion, since it is expectable that comprehension questions about garden-path sentences are harder to answer than comprehension questions about non-ambiguous sentences (Dempsey et al. 2020). Finally, all observations with reading times lower than 100ms and longer than 2000ms were removed, excluding 1% of all observations. The results are qualitatively similar if reading times are unfiltered, or if reading times are log-transformed.

4.2. Results

The mean accuracy of the Control group was 89% (SD = 0.3) and 90% (SD = 0.29) for the Bonus group. Logit models with accuracy as the dependent variable and item order as the predictor were fit for each participant group, revealing that item accuracy increased for the Bonus group during the experiment (β = 0.04, SE = 0.01, z = 3.73, p = 0.0001), but not for the Control group (β = 0.01, SE = 0.006, z = 1.67, p = 0.09). This suggests that the Bonus group participants became better at interpreting the sentences in the experiment correctly, but Control group participants did not. Also, the control group read distractors about 10 ms faster than the Bonus group, per exposure.

In order to avoid the usual convergence problems of Linear Mixed-Effect models, power concerns, and the well-known limitations of frequentist significance testing (Kruschke 2015; Lavine 1999; Sorensen et al. 2016), Bayesian Linear-Mixed effect models were fit, using the BRMS package (Bürkner 2017). The dependent variable was the task-adapted length-corrected residual reading times (TART), with item presentation order, participant group and their interactions as predictors, allowing for the intercept to be adjusted by

participant and item. The model had a flexible threshold and weakly informative priors, and was checked for convergence ($\hat{R} = 1$) after fitting with four chains and 2000 iterations, half of which were the warm-up phase. A significant interaction between participant group and item order (1–16) was found at the spill-over region 8 (see Table 1):

Table 1. Coefficients for the effect on TARTs from the interaction between Control/Bonus reward group and the item order, per region, according to Bayesian Mixed-effect models.

Region	$\beta_{item.order \times group}$	Estimated Error	95% Credible Interval	$P(\beta < 0)$
7 (*was/were*)	0.06	1.05	$[-1.71, 1.78]$	0.48
8 (V)	−1.49	0.96	$[-3.11, 0.08]$	0.94
9 (P)	−0.13	0.68	$[-1.25, 1]$	0.57

Plots illustrating the TART values in regions 7 through 10 are shown in Figure 3.

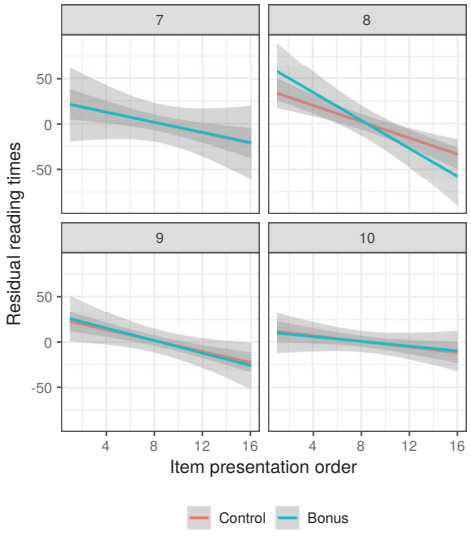


Figure 3. Effect of repeated exposure and reward differential in spill-over region 8.

For completeness, a region-by-region plot with the plain residual reading times is provided in Figure 4. The behaviour of the two groups of participants was generally the same, except that the Bonus group slowed down at region 5 (approaching the critical region), and exhibited greater variability than the Baseline group, which is consistent with participants being more attentive and taking greater care to perform the task.

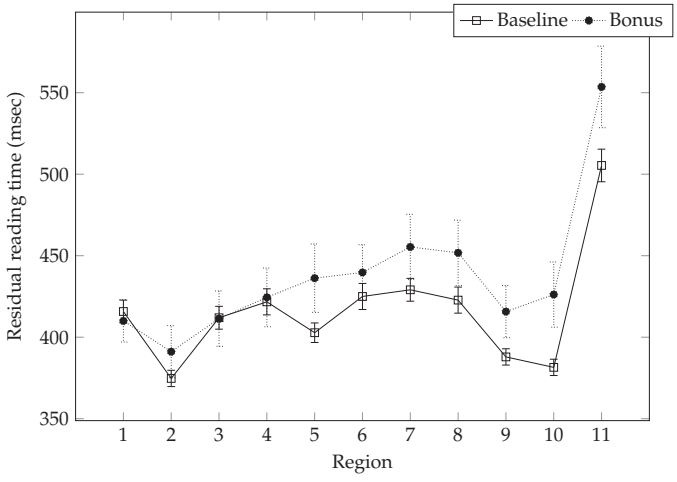


Figure 4. Mean residual reading times for all sentence regions.

4.3. Discussion

The results suggest that participants in the Bonus group used cues in the input to predict the upcoming structure and adapted strategically to the critical items faster than the Control group participants did. Frequency can compound with reward, and speed up sentence processing of complex sentences, in this case, a classic garden-path construction.

It is possible that studies that found null effects in garden-path adaptation (and in island adaptation) were caused by factors that usually are not controlled for: the complexity of the items, their naturalness (norming their non-extracted counterparts would address that), the motivation and focus that participants experience when performing a rather repetitive and artificial task (assigning numbers to sentences, or reading sentences in moving displays), as pointed out by Christianson et al. (2022). To be sure, further research is necessary in order to investigate this matter in more detail, but if it turns out that reward does in fact modulate syntactic adaptation, then a new tool can be added to experimenter's toolkit, which can reduce the chances of null effects caused by low motivation and focus, due to low perceived reward.

5. Conclusions

It is increasingly clear that most island effects are not construction-invariant (Abeillé et al. 2020). Constructions that express assertions tend to yield weaker island effects, for example. Moreover, even in constructions where strong island effects are observed, these are far from categorical. In the present work I have drawn attention to a wide range of factors that likely contribute to that gradience. First, the complexity of the items and the plausibility of the expressed propositions likely plays a role (Hofmeister and Sag 2010). Second, the number of exposures often has an effect, in that it can sometimes cause acceptability ratings to rise. Sometimes that acceptability increase is restricted to the first exposures, sometimes it is not. It is a highly dynamic phenomenon. The acceptability increase instigated by repeated exposure is also selective, in that it does not always affect all sentence types equally. In particular, there is no correlation between sentence acceptability and rate of acceptability change.

The mechanism that drives the amelioration effects remains poorly understood, but extant evidence suggests that speakers are highly sensitive to the items themselves, so that sentences that are excessively complex, or lack semantic plausibility, or require unusual contexts in order to be felicitous in discourse are less likely to improved with repeated

exposure. The amount of exposure also seems to matter, since a number of studies found thresholds after which acceptability increases are observed (Snyder 2021). A survey of the literature on syntactic/task adaptation suggests that syntactic satiation is likely to consist, at least in part, of syntactic adaptation (Fine et al. 2010). This is consistent with the notion that the grammar is gradient and flexible (Francis 2022).

Finally, the present paper puts forth a new factor that can promote adaptation to complex syntactic structures: predicted reward. The underlying mechanism is straightforward: the more motivated and focused the comprehenders are, the faster they can adapt to unusual and complex input, over and above the effect of frequency and task adaptation. This can shed light on why syntactic adaptation – in garden-paths and in certain islands–is not systematically observed in experimental research (Christianson et al. 2022; Kaan and Chun 2018).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1] The traditional view whereby islands are categorical and construction-invariant is still the norm in generative circles, e.g., Željko Bošković (2017), Shafiei and Graf (2020), and many others, though some researchers have more recently come to embrace gradience, like Villata et al. (2019).

[2] Levine (2017, pp. 309–13) points out a possible counterexample to the Conjunct Constraint, shown in (i), in which *succeed* is extracted from the first conjunct VP. But attestations like (j) suggest that (i) may be a special asymmetric disjunction construction in which the reference of the missing subject in the second conjunct is determined anaphorically, rather than via VP coordination.

i. [Succeed] he [[must _] or [be forever shamed]].
j. [[Choose wisely] or [be forever shamed]].

[3] If pragmatics plays a role, this would explain the puzzle raised by Liu et al. (2022) that the matrix verb seems to play a larger role in acceptability than the embedded verb. That is, *What did John say that Mary muttered?* is more acceptable than *What did John mutter that Mary said?* because the latter requires a more unusual context in which John muttered that Mary said something is part of the common ground. This is may be simply a matter of contextualization difficulty (Erteschik-Shir 2006a; Kothari 2008).

[4] All stimuli are available at ...

References

Abeillé, Anne, Olivier Bonami, Danièle Godard, and Jesse Tseng. 2004. The syntax of french n' phrases. In *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven*. Edied by S. Müller. Stanford: CSLI Publications, pp. 6–26.

Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204: 104293. [CrossRef]

Abeillé, Anne, and Rui P. Chaves. 2021. Coordination. In *Head Driven Phrase Structure Grammar*. Empirically Oriented Theoretical Morphology and Syntax. Edited by S. Müller, A. Abeillé, R. D. Borsley, and J.-P. Koenig. Berlin: Language Science Press.

Abrusán, Márta. 2014. *Weak Island Semantics*. Oxford: Oxford University Press.

Altmann, Gerry T. M., and Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73: 247–64. [CrossRef]

Ambridge, Ben, and Adele E. Goldberg. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19: 357–89. [CrossRef]

Anderson, John R. 1990. *The Adaptive Character of Thought*. Lawrence: Erlbaum Associates, Inc.

Arai, Manabu, and Frank Keller. 2013. The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes* 4: 525–60. [CrossRef]

Bondevik, Ingrid. 2018. Investigating the Universality of Adjunct Islands through Formal Acceptability Experiments—A Comparative Study of English and Norwegian. Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway.

Bouma, Gosse, Robert Malouf, and Ivan A. Sag. 2001. Satisfying constraints on extraction andadjunction. *Natural Language & Linguistic Theory* 19: 1–65.

Brandone, Amanda C., Suzanne R. Horwitz, Richard N. Aslin, and Henry M. Wellman. 2014. Infants' goal anticipation during failed and successful reaching actions. *Developmental Science* 17: 23–34. [CrossRef]
Brown, J. M. M., Gisbert Fanselow, Rebecca Hall, and Reinhold Kliegl. 2021. Middle ratings rise regardless of grammatical construction: Testing syntactic variability in a repeated exposure paradigm. *PLoS ONE* 16: e0251280. [CrossRef]
Bürkner, Paul-Christian. 2017. brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software* 80: 1–28. [CrossRef]
Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of cOnscious Experience in Speaking and Writing*. Chicago: University Chicago Press.
Chang, Franklin, Gary S. Dell, and Kathryn Bock. 2006. Becoming syntactic. *Psychological Review* 113: 234–72. [CrossRef]
Chang, Franklin, Marius Janciauskas, and Hartmut Fitz. 2012. Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass* 6: 259–78. [CrossRef]
Chaves, Rui P. 2007. Coordinate Structures—Constraint-Based Syntax-Semantics Processing. Doctoral dissertation, University of Lisbon, Lisboa, Portugal.
Chaves, Rui P. 2012. On the grammar of extraction and coordination. *Natural Language and Linguistic Theory* 30: 465–512. [CrossRef]
Chaves, Rui P. 2013. An expectation-based account of subject islands and parasitism. *Journal of Linguistics* 2: 285–327. [CrossRef]
Chaves, Rui P., and Jeruen E. Dery. 2014. Which subject islands will the acceptability of improve with repeated exposure? In *31st West Coast Conference on Formal Linguistics*. Edited by R. E. Santana-LaBarge. Somerville: Cascadilla Proceedings Project, pp. 96–106.
Chaves, Rui P., and Jeruen E. Dery. 2019. Frequency effects in subject islands. *Journal of Linguistics* 55: 475–521. [CrossRef]
Chaves, Rui P., and Michael T. Putnam. 2021. *Unbounded Dependency Constructions: Theoretical and Experimental Perspectives*. Oxford: Oxford University Press.
Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Edited by R. Freidin, D. Michaels, C. P. Otero, and M. L. Zubizarreta. Cambridge: MIT Press, pp. 133–65.
Christianson, Kiel, Jack Dempsey, Anna Tsiola, and Maria Goldshtein. 2022. *What If They're Just Not that into You (Or Your Experiment)? On Motivation and Psycholinguistics*. Psychology of Learning and Motivation–Advances in Research and Theory. Cambridge: Academic Press Inc. [CrossRef]
Christianson, Kiel, Andrew Hollingworth, John F. Halliwell, and Fernanda Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 4: 368–407. [CrossRef]
Chung, Sandra, and James McCloskey. 1983. On the interpretation of certain island facts in GPSG. *Linguistic Inquiry* 14: 703–14.
Clausen, David R. 2011. Informativity and acceptability of complex subject islands. Paper presented at the 24th Annual CUNY Sentence Processing Conference, Stanford, CA, USA, March 24–26.
Corver, Norbert. 2014. Recursing in dutch. *Natural Language & Linguistic Theory* 32: 423–57.
Creel, Sarah C., Richard N. Aslin, and Michael K. Tanenhaus. 2008. Heeding the voice of experience: The role of talker variation in lexical access. *Cognition* 106: 633–64. [CrossRef]
Culicover, Peter W. 1999. *Syntactic Nuts: Hard Cases in Syntax*. Volume One of Foundations of Syntax. Oxford: Oxford University Press.
Culicover, Peter W. 2013. *Grammar & Complexity: Language at the Interface of Competence and Performance*. Oxford: Oxford University Press.
Culicover, Peter W., and Susanne Winkler. 2022. Parasitic gaps aren't parasitic, or, the case of the uninvited guest. *The Linguistic Review* 39: 1–35. [CrossRef]
Davies, Mark. 2008. The Corpus of Contemporary American English: 450 Million Words, 1990-Present. Available online: http://corpus.byu.edu/coca/ (accessed on 3 March 2022).
Deane, Paul. 1991. Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics* 2: 1–63. [CrossRef]
Degen, Judith, and Judith Tonhauser. 2022. Are there factive predicates? An empirical investigation. *Language forthcoming*.
DeLong, Katherine A., Thomas P. Urbach, and Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8: 1117–21. [CrossRef] [PubMed]
Dempsey, Jack, Qiawen Liu, and Kiel Christianson. 2020. Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 46: 1906–21. [CrossRef] [PubMed]
Do, Monica L., and Elsi Kaiser. 2017. The relationship between syntactic satiation and syntactic priming: A first look. *Frontiers in Psychology* 8: 1851. [CrossRef] [PubMed]
Emonds, Joseph E., and Jan Terje Faarlund. 2014. *English: The Language of the Vikings*. Olomouc: Palacký University.
Engdahl, Elisabet. 1983. Parasitic gaps. *Linguistics and Philosophy* 6: 3–34. [CrossRef]
Epstein, Dwaine. 2013. *Lee Marvin: Point Blank*. Tucson: Schaffner Press.
Erteschik-Shir, Nomi. 1977. *On the Nature of Island Constraints*. Bloomington: Indiana University Linguistics Club.
Erteschik-Shir, Nomi. 2006a. Bridge phenomena. In *The Blackwell Companion to Syntax Volumes*. Edited by M. Everaert, H. V. Riemsdijk, R. Goedemans, and B. Hollebrandse. Oxford: Blackwell, vol. 5, pp. 284–94.
Erteschik-Shir, Nomi. 2006b. What's what? In *Gradience in Grammar*. Edited by G. Fanselow, C. Fery, M. Schlesewsky, and R. Vogel. Oxford: Oxford University Press, pp. 317–35.
Erteschik-Shir, Nomi, and Shalom Lappin. 1979. Dominance and the functional explanation of island phenomena. *Theoretical Linguistics* 6: 41–86. [CrossRef]

Farmer, Thomas A., Alex B. Fine, Shaorong Yan, Spyridoula Cheimariou, and Florian T. Jaeger. 2014. Error-driven adaptation of higher-level expectations during natural reading. Paper presented at the 36th Annual Meeting of the Cognitive Science Society, Quebec City, QC, Canada, July 23–26. Austin: Cognitive Science Society.
Ferreira, Fernanda, and John M. Henderson. 1990. Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16: 555–68. [CrossRef]
Ferreira, Fernanda, and John M. Henderson. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 31: 725–45. [CrossRef]
Fine, Alex, and Florian T. Jaeger. 2016. The role of verb repetition in cumulative structural priming in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42: 1362–76. [CrossRef]
Fine, Alex B., Florian T. Jaeger, Thomas A. Farmer, and Ting Qian. 2013. Rapid expectation adaptation during syntactic comprehension. *PLoS ONE* 8: e77661. [CrossRef]
Fine, Alex B., and T. Florian Jaeger. 2013. Evidence for implicit learning in syntactic comprehension. *Cognitive Science* 37: 578–91. [CrossRef]
Fine, Alex B., Ting Qian, Florian T. Jaeger, and Robert A. Jacobs. 2010. Syntactic adaptation in language comprehension. Paper presented at the ACL Workshop on Cognitive Modeling and Computational Linguistics, Uppsala, Sweden, July 15, pp. 18–26.
Francis, Elaine. 2022. *Gradient Acceptability and Linguistic Theory*. Oxford: Oxford University Press.
Francom, Jerid. 2009. Experimental Syntax: Exploring the Effect of Repeated Exposure to Anomalous Syntactic Structure–Evidence from Rating and Reading Tasks. Ph.D. thesis, University of Arizona, Tucson, AZ, USA.
Frazier, Lyn, and Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14: 178–210. [CrossRef]
Gahl, Susanne, Dan Jurafsky, and Douglas Roland. 2004. Verb subcategorization frequencies: American english corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, and Computers* 36: 432–43. [CrossRef] [PubMed]
Gibson, Edward. 2006. The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language* 54: 363–38. [CrossRef]
Gibson, Edward, Barbara Hemforth, Elodie Winckel, and Anne Abeillé. 2021. Acceptability of extraction out of adjuncts depends on discourse factors. Paper presented at the 34th Annual CUNY Conference on Human Sentence Processing, Philadelphia, PA, USA, March 4–6.
Gigerenzer, Gerd, Peter M. Todd, and the ABC Research Group. 1999. *Simple Heuristics that Make us Smart*. Oxford: Oxford University Press.
Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
Goldberg, Adele E. 2013. Backgrounded constituents cannot be extracted. In *Experimental Syntax and Island Effects*. Edited by J. Sprouse and N. Hornstein. Cambridge: Cambridge University Press, pp. 221–38.
Goodall, Grant. 2011. Syntactic satiation and the inversion effect in English and Spanish *wh*- questions. *Syntax* 14: 29–47. [CrossRef]
Grosu, Alexander. 1973. On the nonunitary nature of the coordinate structure constraint. *Linguistic Inquiry* 4: 88–92.
Heathcote, Andrew, Scott Brown, and D. J. K. Mewhort. 2000. The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review* 7: 185–207.
Hegarty, Michael. 1990. On adjunct extraction from complements. In *MIT Working Papers in Linguistics 13*. Edited by L. L.Cheng and H. Demirdash. Cambridge: MITWPL, pp. 101–24.
Hiramatsu, Kazuko. 2000. Accessing Linguistic Competence: Evidence from Children's and Adults? Acceptability Judgments. Ph.D. thesis, University of Connecticut, Storrs, CT, USA.
Hofmeister, Philip. 2011. Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes* 3: 376–405. [CrossRef]
Hofmeister, Philip. 2015. Experience-driven acceptability effects. *Unpublished manuscript*.
Hofmeister, Philip, and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415. [CrossRef]
Huang, Cheng-Teh James. 1982. Logical Relations in Chinese and the Theory of Grammar. Ph.D. thesis, MIT, Cambridge, MA, USA.
Huddleston, Rodney D., Geoffrey Pullum, and Peter Peterson. 2002. *The Cambridge Grammar of the English Language*. Chapter 12: Relative clause constructions and unbounded dependencies. Cambridge: Cambridge University Press, pp. 1033–96.
Jacob, Gunnar, and Claudia Felser. 2016. Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology* 69: 907–25. [CrossRef]
Jaeger, Florian T., Zachary Burchill, and Wednesday Bushong. 2018. Strong Evidence for Expectation Adaptation during Language Understanding, Not a Replication Failure. A Reply to Harrington Stack, James, and Watson. Available online: https://osf.io/4vxyp/ (accessed on 3 March 2022).
Jiménez–Fernández, Ángel L. 2009. On the composite nature of subject islands: A phase-based approach. *SKY Journal of Linguistics* 22: 91–138.
Just, Marcel Adam, Patricia A. Carpenter, and Jacqueline D. Wooley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111: 228–38. [CrossRef]
Kaan, Edith, and Eunjin Chun. 2018. Syntactic adaptation. In *The Psychology of Learning and Motivation: Current Topics in Language*. Edited by K. D. Federmeier and D. G. Watson. Amsterdam: Elsevier; Cambridge: Academic Press, pp. 85–116.

Kamide, Yuki, and Don C. Mitchell. 1997. Relative clause attachment: Non-determinism in Japanese parsing. *Journal of Psycholinguistic Research* 26: 247–54. [CrossRef]
Kayne, Richard. 1981. ECP Extensions. *Linguistic Inquiry* 12: 93–133.
Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford: CSLI Publications.
Kim, Ilkyu. 2021. A note on the factorial definition of island effects. *Second Language Research* 57: 211–23. [CrossRef]
Kluender, Robert. 1998. On the distinction between strong islands and weak islands: A processing perspective. In *Syntax and Semantics 29: The Limits of Syntax*. Edited by P. W. Culicover and L. McNally. New York: Academic Press, pp. 241–79.
Kluender, Robert. 2004. Are subject islands subject to a processing account? In *Proceedings of the WCCFL 23*. Editd by B. Schmeiser, V. Chand, A. Kelleher and A. Rodriguez. Somerville: Cascadilla Press, pp. 101–25.
Kohrt, Annika, Trey Sorense, and Dustin Chacón. 2018a. Syntax predicts, semantics revises: Filler-gap dependencies into adjunct clauses. Paper presented at the 31st Annual CUNY Sentence Processing Conference, University of California, Davis, CA, USA, March 15–17.
Kohrt, Annika, Trey Sorensen, and Dustin A. Chacón. 2018b. The real-time status of semantic exceptions to the adjunct island constraint. Paper presented at the WECOL 2018: Western Conference on Linguistics, Fresno, CA, USA, November 30–December 2.
Kohrt, Annika, Trey Sorensen, Peter O'Neill, and Dustin A. Chacón. 2020. Inactive gap formation: An erp study on the processing of extraction from adjunct clauses. Paper presented at the LSA 2020 Annual Meeting, New Orleans, LA, USA, January 2–5.
Kothari, Anubha. 2008. Frequency-based expectations and context influence bridge quality. In *Proceedings of WECOL 2008: Western Conference on Linguistics*. Edited by M. Grosvald and D. Soares. Fresno: California State University, pp. 136–49.
Kruschke, John. 2015. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd ed. London: Academic Press.
Kubota, Yusuke, and Jungmee Lee. 2015. The coordinate structure constraint as a discourse-oriented principle: Further evidence from japanese and korean. *Language* 91: 642–75. [CrossRef]
Kuno, Susumo. 1972. Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry* 3: 299–320.
Kuno, Susumu. 1987. *Functional Syntax–Anaphora, Discourse and Empathy*. Chicago and London: The University of Chicago Press.
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2018. Investigating variation in island effects. *Natural Language & Linguistic Theory* 36: 743–79.
Kush, Dave, Terje Lohndal, and Jon Sprouse. 2019. On the island sensitivity of topicalization in norwegian: An experimental investigation. *Language* 95: 393–420. [CrossRef]
Kush, David, Akira Omaki, and Norbert Hornstein. 2013. Microvariation in islands? In *Experimental Syntax and Island Effects*. Edited by J. Sprouse and N. Hornstein. Cambridge: Cambridge University Press, pp. 239–64.
Kutas, Marta, and Steven A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 5947: 161–63. [CrossRef] [PubMed]
Lakoff, George. 1986. Frame Semantic Control of the Coordinate Structure Constraint. In *Papers from the 22nd Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.
Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.
Lau, Ellen, Clare Stroud, Silke Plesch, and Colin Phillips. 2006. The role of structural prediction in rapid syntactic analysis. *Brain and Language* 1: 74–88. [CrossRef]
Lavine, Michael. 1999. What is bayesian statistics and why everything else is wrong. *The Journal of Undergraduate Mathematics and Its Applications* 20: 165–74.
Legenstein, Robert, Dejan Pecevski, and Wolfgang Maass. 2008. A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. *PLoS Computational Biology* 4: e1000180. [CrossRef]
Levine, Robert D. 2017. *Syntactic Analysis: An HPSG-Based Approach*. Cambridge: Cambridge University Press.
Levine, Robert D., and Thomas E. Hukari. 2006. *The Unity of Unbounded Dependency Constructions*. Stanford: CSLI Publications.
Levine, Robert D., and Ivan A. Sag. 2003. Some empirical issues in the grammar of extraction. In *Proceedings of the HPSG-2003 Conference, Michigan State University, East Lansing*. Edited by S. Müller. Stanford: CSLI Publications, pp. 236–56.
Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 3: 1126–77. [CrossRef]
Levy, Roger, Evelina Fedorenko, Mara Breen, and Ted Gibson. 2012. The processing of extraposed structures in English. *Cognition* 12: 12–36. [CrossRef]
Levy, Roger, and Frank Keller. 2013. Expectation and locality effects in German verb-final structures. *Journal of Memory and Language* 2: 199–222. [CrossRef]
Liu, Yingtong, Rachel Ryskin, Richard Futrell, and Edward Gibson. 2019. Verb frequency explains the unacceptability of factive and manner-of-speaking islands in english. Paper presented at the 41st Annual Meeting of the Cognitive Science Society, Montreal, QC, Canada, July 24–27.
Liu, Yingtong, Rachel Ryskin, Richard Futrell, and Edward Gibson. 2022. A verb-frame frequency account of constraints on long-distance dependencies in english. *Cognition* 222: 104902. [CrossRef]
Lu, Jiayi, Daniel Lassiter, and Judith Degen. 2021. Syntactic satiation is driven by speaker-specific adaptation. *Proc. Annu. Meet. Cogn. Sci. Soc.* 43: 1493–99.
MacDonald, Maryellen C., Neal J. Pearlmutter, and Mark S. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101: 676–703. [CrossRef] [PubMed]

Malone, Avery, and Gail Mauner. 2018. What do readers adapt to in syntactic adaptation? Paper presented at the 31st Annual CUNY Sentence Processing Conference, Davis, CA, USA, March 15–17.
Malone, Avery, and Gail Mauner. 2020. Syntactic adaptation for reduced relative clauses is not reducible to task adaptation. Paper presented at the Poster Session, 33rd CUNY Conference on Human Sentence Processing, Amherst, MA, USA, March 19–21.
McCawley, James D. 1981. The syntax and semantics of english relative clauses. *Lingua* 53: 99–149. [CrossRef]
Metzing, Charles, and Susan E. Brennan. 2003. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language* 49: 201–13. [CrossRef]
Michaelis, Laura, and Hartwell Francis. 2007. Lexical subjects and the conflation strategy. In *Topics in the Grammar-Pragmatics Interface: Papers in honor of Jeanette K. Gundel*. Edited by N. Hedberg and R. Zacharski. Amsterdam: John Benjamins.
Müller, Christiane. 2017. Extraction from adjunct islands in Swedish. *Norsk Lingvistisk Tidsskrift* 35: 67–85.
Müller, Christiane, and Clara Ulrich Eggers. 2022. Island extractions in the wild: A corpus study of adjunct and relative clause islands in danish and english. *Languages* 7: 125.
Na, Younghee, and Geoffrey Huck. 1992. On extracting from asymmetrical structures. In *The Joy of Grammar*. Edited by D. Brentari, G. N. Larson, and L. A. MacLeod. Amsterdam: John Benjamins, pp. 251–74.
Newell, Allen, and Herbert A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs: Prentice-Hall.
Noppeney, Uta, and Cathy J Price. 2004. An fmri study of syntactic adaptation. *Journal of Cognitive Neuroscience* 16: 702–13. [CrossRef]
Oshima, David Y. 2007. On factive islands: Pragmatic anomaly vs. pragmatic infelicity. Paper presented at the JSAI'06 20th Annual Conference on New Frontiers in Artificial Intelligence, Tokyo, Japan, June 5–9. Berlin/Heidelberg: Springer, pp. 147–61.
Otmakhova, Nonna A., and John E. Lisman. 1996. D1/d5 dopamine receptor activation increases the magnitude of early long-term potentiation at ca1 hippocampal synapses. *Journal of Neuroscience* 16: 7478–86. [CrossRef]
Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford CSLI.
Prasad, Grusha, and Tal Linzen. 2019. Reassessing the evidence for syntactic adaptation from self-paced reading studies. Paper presented at the Poster Session, 32nd CUNY Conference on Human Sentence Processing, Boulder, CO, USA, March 31.
Prasad, Grusha, and Tal Linzen. 2021. Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory and Cognition* 47: 1156–72. [CrossRef]
Reynolds, John N., Brian I. Hyland, and Jeffrey R. Wickens. 2001. A cellular mechanism of reward-related learning. *Nature* 413: 67–70. [CrossRef]
Roland, Douglas, Frederic Dick, and Jeffrey L. Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57: 348–379. [CrossRef]
Ross, John R. 1967. Constraints on Variables in Syntax. Ph.D. dissertation, MIT, Cambridge, MA, USA. [Published in 1986 as *Infinite Syntax!* Norwood: Ablex Publishing].
Sauerland, Uli, and Paul D. Elbourne. 2002. Total reconstruction, pf movement, and derivational order. *Linguistic Inquiry* 33: 283–319. [CrossRef]
Schultz, Wolfram. 1998. Predictive reward signal of dopamine neurons. *Journal of Neurophysiology* 80: 1–27. [CrossRef] [PubMed]
Schwarz, Bernhard, and Alexandra Simonenko. 2018. Factive islands and meaning-driven unnaceptability. *Natural Language Semantics* 26: 253–79. [CrossRef]
Shafiei, Nazila, and Thomas Graf. 2020. The subregular complexity of syntactic islands. In *Proceedings of the Society for Computation in Linguistics*. Amherst: University of Massachusetts Amherst, vol. 3.
Sikos, Les, H. Martin, Laura Fitzgerald, and D. Grodner. 2016. Memory-based limits on surprisal-based syntactic adaptation. Paper presented at the 29th Annual CUNY Conference of Human Sentence Processing, Gainesville, FL, USA, March 3–5.
Snyder, William. 1994. A psycholinguistic investigation of weak crossover, islands, and syntactic satiation effects: Implications for distinguishing competence from performance. Paper presented at the 7th CUNY Conference of Human Sentence Processing, New York, NY, USA, March 17–19.
Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31: 575–82. [CrossRef]
Snyder, William. 2017. On the nature of syntactic satiation. *Unpublished manuscript*.
Snyder, William. 2021. Satiation. In *The Cambridge Handbook of Experimental Syntax*. Edited by G. Goodall. Cambridge: Cambridge University Press.
Sorensen, Tanner, Sven Hohenstein, and Shravan Vasishth. 2016. Bayesian linear mixed models using stan: A tutorial for psychologists, linguists, and cognitive scientists. *Quantitative Methods for Psychology* 12: 175–200. [CrossRef]
Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1: 118–29. [CrossRef]
Sprouse, Jon. 2009. Revisiting satiation: Evidence for an equalization response strategy. *Linguistic Inquiry* 40: 329–41. [CrossRef]
Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in english and italian. *Natural Language & Linguistic Theory* 34: 307–44.
Sprouse, Jon, Matt Wagers, and Colin Phillips. 2012. A test of the relation between working memory capacity and syntactic island effects. *Language* 88: 82–123. [CrossRef]
Stack, Caoimhe M. Harrington, Ariel N. James, and Duane G. Watson. 2018. A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition* 46: 864–77.
Staub, Adrian. 2007. The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33: 550–69. [CrossRef] [PubMed]

Staub, Adrian, and Charles Clifton, Jr. 2006. Syntactic prediction in language comprehension: Evidence from *either... or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32: 425–36. [CrossRef] [PubMed]
Stromswold, Karin. 1986. *Syntactic Satiation*. Cambridge: MIT. *Unpublished Ms*.
Szabolcsi, Anna, and Terje Lohndal. 2017. Strong vs. weak islands. In *The Wiley Blackwell Companion to Syntax*, 2nd ed. Hoboken: John Wiley & Sons, Inc., pp. 1–51.
Szabolcsi, Anna, and Frans Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics* 1: 235–84. [CrossRef]
Tabor, Whitney, C. Juliano, and Michael K. Tanenhaus. 1997. Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes* 2–3: 211–71. [CrossRef]
Takami, Ken-Ichi. 1988. Preposition stranding: Arguments against syntactic analyses and an alternative functional explanation. *Lingua* 76: 299–335. [CrossRef]
Tonhauser, Judith, David Beaver, and Judith Degen. 2018. How projective is projective content? gradience in projectivity and at-issueness. *Journal of Semantics* 35: 495–542. [CrossRef]
Traxler, Matthew J. 2015. Priming of early closure: Evidence for the lexical boost during. *Language, Cognition and Neuroscince* 30: 478–90. [CrossRef]
Traxler, Matthew J., and Martin J. Pickering. 2005. Syntactic priming in comprehension: Evidence from eye-movements. Paper presented at the 18th Annual CUNY Conference on Human Sentence Processing, Tucson, AZ, USA, March 31–April 2.
Trueswell, John C., and Michael K. Tanenhaus. 1994. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33: 258–318. [CrossRef]
Truswell, Robert. 2011. *Events, Phrases and Questions*. Oxford: Oxford University Press.
Uriagereka, Juan. 1988. On Government. Ph. D. thesis, University of Connnecticut, Storrs, CT, USA.
Van Valin, Robert D., Jr. 1986. Pragmatics, island phenomena, and linguistic competence. In *Papers from the Parasession on Pragmatics and Grammatical Theory*. Edited by A. Farley, P. Farley and K.-E. McCullough. Chicago: Chicago Linguistic Society, vol. 22, pp. 223–33.
Villata, Sandra, Jon Sprouse, and Whitney Tabor. 2019. Modeling ungrammaticality: A self-organizing model of islands. Paper presented at the 41st Annual Meeting of the Cognitive Science Society, Montreal, QC, Canada, July 24–27, pp. 1178–84.
Vincent, Jake Wayne. 2021. Extraction from Relative Clauses: An Experimental Investigation into Variable Island Effects in English. Ph.D. dissertation, University of California, Santa Cruz, CA, USA.
Vincent, Jake W., Ivy Sichel, and Matthew W. Wagers. 2022. Extraction from english rcs and cross-linguistic similarities in the environments that facilitate extraction. *Languages* 7: 117. [CrossRef]
Wells, Justine, Morten H. Christiansen, David S. Race, Daniel J. Acheson, and Maryellen C. MacDonald. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58: 250–71. [CrossRef]
Xu, Fei, and Tamar Kushnir. 2013. Infants are rational constructivist learners. *Current Directions in Psychological Science* 22: 28–32. [CrossRef]
Zehr, Jeremy, and Florian Schwarz. 2018. Penncontroller for Internet Based Experiments (Ibex). Ms. Available online: https://osf.io/md832/ (accessed on 3 March 2022).
Željko Bošković. 2017. On unifying the coordinate structure constraint and the adjunct condition. In *Syntactic Architecture and Its Consequences II: Between Syntax and Morphology (Open Generative Syntax 10)*. Edited by A. Bárány, T. Biberauer, J. Douglas, and S. Vikner. Berlin: Language Sciences Press.

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Languages Editorial Office
E-mail: languages@mdpi.com
www.mdpi.com/journal/languages

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34

www.mdpi.com

ISBN 978-3-0365-6317-6